

# Emergence of a Giant Component in Random Site Subgraphs of a $d$ -Dimensional Hamming Torus

David Sivakoff\*

Department of Mathematics, University of California, Davis.

April 9, 2010

## Abstract

The  $d$ -dimensional Hamming torus is the graph whose vertices are all of the integer points inside an  $a_1 n \times a_2 n \times \cdots \times a_d n$  box in  $\mathbb{R}^d$  (for constants  $a_1, \dots, a_d > 0$ ), and whose edges connect all vertices within Hamming distance one. We study the size of the largest connected component of the subgraph generated by independently removing each vertex of the Hamming torus with probability  $1 - p$ . We show that if  $p = \frac{\lambda}{n}$ , then there exists  $\lambda_c > 0$ , which is the positive root of a degree  $d$  polynomial whose coefficients depend on  $a_1, \dots, a_d$ , such that for  $\lambda < \lambda_c$  the largest component has  $O(\log n)$  vertices (a.s. as  $n \rightarrow \infty$ ), and for  $\lambda > \lambda_c$  the largest component has  $(1 - q)\lambda \left(\prod_i a_i\right) n^{d-1} + o(n^{d-1})$  vertices and the second largest component has  $O(\log n)$  vertices (a.s.). An implicit formula for  $q < 1$  is also given. Surprisingly, the value of  $\lambda_c$  that we find is distinct from the critical value for the emergence of a giant component in the random edge subgraph of the Hamming torus. Additionally, we show that if  $p = \frac{c \log n}{n}$ , then when  $c < \frac{d-1}{\sum a_i}$  the site subgraph of the Hamming torus is not connected, and when  $c > \frac{d-1}{\sum a_i}$  the subgraph is connected (a.s.). We also show that the subgraph is connected precisely when it contains no isolated vertices.

## 1 Introduction

Erdős and Rényi studied random subgraphs of the complete graph in [9] (for an account of their results, see [8]). In the Erdős-Rényi model, a random subgraph of the complete graph on  $n$  nodes is obtained by independently deciding whether to remove each edge with probability  $(1 - p)$  or keep it with probability  $p$ . Erdős and Rényi studied the size of the largest connected component in the random subgraph under the scaling  $p = \lambda/n$ , where  $\lambda$  is a constant parameter. They found that the size of the largest connected component is asymptotically almost surely  $O(\log n)$  for  $\lambda < 1$ , and  $cn + o(n)$  for  $c = c(\lambda) > 0$  when  $\lambda > 1$ . A sequence of events  $E_n$  is said to occur asymptotically almost surely (a.s.) if  $\mathbb{P}(E_n) \rightarrow 1$  as  $n \rightarrow \infty$ . Thus, at  $\lambda = 1$ , the Erdős-Rényi model undergoes a transition from having small components to having a giant connected component.

Random subgraphs generated by randomly removing edges independently with probability  $(1 - p)$  have since been extensively studied on other graphs  $G = (V, E)$ , and we will refer to random subgraphs generated in this manner as random edge subgraphs. Alternatively, random subgraphs

---

\*Dissertation research at UC Davis supported in part by NSF VIGRE Grant No. DMS-0636297, and NSF Grant Nos. DMS-0805970 and DMS-0505734.

of  $G$  can be obtained by independently removing vertices (and any edges incident to those vertices) with probability  $(1 - p)$ ; we will refer to these as random site subgraphs.

The  $n$ -dimensional hypercube is an example of a graph for which both random edge and site subgraphs have been studied. Ajtai et al. [1] studied random edge subgraphs of the  $n$ -dimensional hypercube:  $V = \{0, 1\}^n$ ,  $E = \{(x, y) \in V \times V : d(x, y) = 1\}$ , where  $d(x, y)$  is the Hamming distance between  $x$  and  $y$  (the number of coordinates in which they differ). For the scaling  $p = \lambda/n$ , they proved that all components of the random edge subgraph have  $O(n)$  vertices when  $\lambda < 1$  (a.a.s.), and for  $\lambda > 1$  there is a component with  $c2^n + o(2^n)$  vertices (a.a.s.) for  $c = c(\lambda) > 0$ . Bollobás et al. [5] found similar behavior for the random site subgraph of the  $n$ -dimensional hypercube. Under the same scaling, they proved that for  $\lambda < 1$  all components have at most  $O(n)$  vertices (a.a.s.), and for  $\lambda > 1$  there is a component with  $cn^{-1}2^n + o(n^{-1}2^n)$  vertices for  $c = c(\lambda) > 0$  (a.a.s.). Thus, both random subgraph models have the same threshold under the scaling  $p = \lambda/n$  for the  $n$ -dimensional hypercube. In fact, the value of  $c(\lambda)$  is the same for both models, so even the proportion of vertices in the giant component above the threshold is the same.

The intuition behind each of these results is that random edge subgraphs of the hypercube and the complete graph and random site subgraphs of the hypercube look locally tree-like. So, the connected component associated with a fixed vertex can be compared to a branching process with a Binomial( $D, p$ ) offspring distribution, where  $D$  is the degree of each vertex. A giant connected component exists precisely when the branching process has positive survival probability, which is when the expected number of offspring per individual ( $Dp = \lambda$ ) exceeds 1. Borgs et al. [6, 7] showed that, for a wide variety of finite transitive graphs, the threshold for the existence of a giant component in the random edge subgraph should be at  $Dp = 1$ , though they focus on behavior in the critical window, and do not give a lower bound for the size of the supercritical giant component. The case of random site subgraphs is less clear — just consider the complete graph. All random site subgraphs of the complete graph are connected, thus the threshold for the existence of a giant component is 0. This is a trivial case, but suggests that the random edge and site subgraphs will behave differently when the edge density of  $G$  is closer to that of the complete graph than to the hypercube.

Random site subgraphs are particularly relevant to problems that arise from the theory of genetic fitness landscapes. Fitness landscapes are a tool used in theoretical evolutionary biology to model speciation under various conditions [10]. A fitness landscape consists of a genotype space and a function that maps each combination of genes to a fitness level in the interval  $[0, 1]$ , which can be interpreted as the probability of an individual with that genotype surviving to reproduce. For example, the genotype space for a haploid genome with  $n$  diallelic loci (two gene alternatives at each position along the genome) can be represented by the  $n$ -dimensional hypercube, where each vertex represents a genotype and edges represent single-gene mutations. Each genotype is then assigned a fitness level of 1 with probability  $p$  or 0 with probability  $(1 - p)$ , independent of all other genotypes — this is equivalent to taking a random site subgraph. For this model, provided  $n$  is sufficiently large, only a small proportion of genotypes need to be viable ( $p > 1/n$ ) for a significant number of viable genotypes to be accessible via single-gene mutations.

The random graph model that we are interested in comes from a genotype space for a haploid genome with  $d$  loci, and  $a_i n$  possible alleles at the  $i^{\text{th}}$  locus. Here  $d$  is fixed, and the number of alleles at each locus is assumed to be large. The resulting graph is a Hamming torus in  $d$  dimensions:

$$V = \{(x_1, \dots, x_d) \in \mathbb{Z}^d : 1 \leq x_i \leq a_i n, i = 1, \dots, d\}$$

$$E = \{(x, y) \in V \times V : d(x, y) = 1\},$$

where  $d(x, y)$  is the Hamming distance between  $x$  and  $y$ . For  $d = 1$  this is the complete graph on  $a_1 n$  vertices, and for  $d = 2$  this is the rook graph on an  $a_1 n$  by  $a_2 n$  chessboard. As before, fitness levels

of 1 are assigned independently with probability  $p$ , and fitness of 0 is otherwise inferred. We say that sites with a fitness of 1 are *occupied*, while those with a fitness of 0 are *removed* or unoccupied.

This model is amenable to comparison with a multitype branching process. Consider a multitype branching process with  $d$  types in which an individual of type  $i$  has a  $\text{Binomial}(a_j n, \frac{\lambda}{n})$  number of offspring of type  $j$  for  $j \neq i$  and zero offspring of type  $i$ . This branching process has nonzero survival probability iff the largest eigenvalue of the matrix of expected progeny,  $\mathbf{M}_\lambda$ , is strictly larger than one [2].

$$\mathbf{M}_\lambda = \begin{pmatrix} 0 & \lambda a_2 & \lambda a_3 & \cdots & \lambda a_d \\ \lambda a_1 & 0 & \lambda a_3 & \cdots & \lambda a_d \\ \lambda a_1 & \lambda a_2 & 0 & & \vdots \\ \vdots & \vdots & & \ddots & \lambda a_d \\ \lambda a_1 & \lambda a_2 & \cdots & \lambda a_{d-1} & 0 \end{pmatrix}$$

where the entry in position  $(i, j)$  is the expected number of type  $j$  offspring to which a type  $i$  individual will give birth. This branching process is a good local approximation for clusters in the random site subgraph of the Hamming torus when  $p = \frac{\lambda}{n}$ , which motivates the following theorems.

**Theorem 1.1.** *Fix  $d \geq 2$  and  $a_1, \dots, a_d > 0$ , and let  $\lambda_c$  be the unique positive solution to  $\det(\mathbf{M}_{\lambda_c} - \mathbf{I}) = 0$ . If  $\lambda < \lambda_c$  and  $p = \frac{\lambda}{n}$  then the largest connected component of the random site subgraph of the Hamming torus is a.a.s.  $O(\log n)$ .*

**Theorem 1.2.** *Fix  $d \geq 2$  and  $a_1, \dots, a_d > 0$ , and let  $\lambda_c$  be the unique positive solution to  $\det(\mathbf{M}_{\lambda_c} - \mathbf{I}) = 0$ . If  $\lambda > \lambda_c$  and  $p = \frac{\lambda}{n}$  then the size of the largest connected component of the random site subgraph of the Hamming torus is a.a.s.  $(1 - q)\lambda(\prod_i a_i) n^{d-1} + o(n^{d-1})$ . Furthermore, the second largest component has a.a.s.  $O(\log n)$  vertices.*

For Theorem 1.2,  $(1 - q) > 0$  is defined precisely in Section 3.4, and in both theorems  $\mathbf{I}$  is the  $d \times d$  identity matrix. Note that the total number of vertices in the random site subgraph of the Hamming torus is asymptotically the expected number,  $p|V| = \lambda(\prod_i a_i) n^{d-1}$ , so  $(1 - q)$  is the proportion of vertices remaining that are in the giant component. The critical value,  $\lambda_c$ , can be defined equivalently as the root of a degree  $d$  polynomial due to the following lemma.

**Lemma 1.3.**

$$\det(\mathbf{M}_\lambda - \mathbf{I}) = (-1)^d \left[ 1 - \sum_{\ell=2}^d (\ell - 1) \lambda^\ell \sum_{\substack{S \subset \{1, \dots, d\} \\ |S| = \ell}} \prod_{i \in S} a_i \right].$$

Random *edge* subgraphs of the Hamming torus were considered by Borgs et al. [6, 7], who indicated that the threshold for the emergence of a giant component in the random edge subgraph should be  $1/(a_1 + \dots + a_d)$ , though they do not provide a lower bound for the size of the largest component above the threshold. This lower bound was later proven by van der Hofstad and Luczak [12] in the case  $d = 2$  and  $a_1 = a_2 = 1$ , demonstrating that the threshold occurs at  $1/2$ . When  $d = 2$ , the threshold for the random site subgraph is  $\lambda_c = 1/\sqrt{a_1 a_2}$ , so the two processes clearly differ.

Multitype branching processes have also been employed in the analysis of *inhomogeneous* random edge subgraphs of the complete graph by Bollobás et al. [4]. In the inhomogeneous model, the probability of retaining the edge between vertices  $i$  and  $j$  is  $p_{ij}$ , and the  $p_{ij}$ 's may not be equal, however the inclusion or exclusion of each edge still occurs independently of all other edges. Since

the number of neighbors that a vertex has depends on the probabilities of the edges incident to that vertex, the inhomogeneous model looks locally like a multitype branching process — each vertex clearly has a different offspring distribution, and is thus of a different “type”. Bollobás et al. [4] proved that, under certain conditions on the values of the  $p_{ij}$ ’s (which exclude the case of the Hamming torus), a giant component exists precisely when the corresponding multitype branching process survives. In the case of random site subgraphs of the Hamming torus, it is somewhat surprising that a connection to multitype branching processes arises, since each vertex sees the same distribution of neighbors. In this model, the similarity to a multitype branching process is due to the dependencies between edges, and not the inhomogeneity of the edge probabilities as in [4].

In Section 2 we prove Theorem 1.1 by coupling a process of revealing the vertices in a connected component of the random site subgraph with a multitype branching process. In Section 3 we prove Theorem 1.2 in four steps. The first step, which is in Section 3.2, is to show that the process of revealing the vertices in a connected component of the random site subgraph will either terminate before discovering  $O(\log n)$  occupied vertices, or will reach size  $m = \Theta(\log n)$  with high probability. The second step, in Section 3.1, is to show that if the process of revealing vertices reaches size  $m$ , then it can be coupled with a lower bounding branching process, which will reach size  $n^{d-4/3}$  with high probability. The third step, in Section 3.3, is to show that any two component-discovering processes (started from two different vertices in  $V$ ) that reach size  $n^{d-4/3}$  will join together with high probability. The final step, in Section 3.4, is to show that the proportion of vertices in components of size  $O(\log n)$  converges to  $q$  in probability. The reason we reverse the order in which the first two steps are presented is because  $m$  needs to be defined before proving the first step, but this definition is motivated by the second step. In Section 4 we will prove the following three theorems regarding the connectivity of the random site subgraph.

**Theorem 1.4.** *Let  $c < \frac{d-1}{\sum a_i}$ . If  $p = p(n) \leq \frac{c \log n}{n}$  and  $p = \omega(n^{-d})$  then the random site subgraph of the Hamming torus contains isolated vertices, and is thus not connected (a.a.s.).*

**Theorem 1.5.** *Let  $c > \frac{d-1}{\sum a_i}$ . If  $p = p(n) \geq \frac{c \log n}{n}$  then the random site subgraph of the Hamming torus is connected (a.a.s.).*

**Theorem 1.6.** *Fix  $a_1 \geq a_2 \geq \dots \geq a_d$ , and let  $c > \frac{d-1}{2 \sum_{i=2}^d a_i + a_1}$ . If  $p = \frac{c \log n}{n}$  then every vertex in the random site subgraph of the Hamming torus is either isolated or belongs to the giant component (a.a.s.).*

These three theorems together imply that, with probability approaching one, the random site subgraph of the Hamming torus is connected if and only if it contains no isolated vertices (except in the trivial case  $p \asymp n^{-d}$ , where the subgraph may consist of just a single occupied vertex with positive probability).

## 2 Subcritical behavior

Theorem 1.1 comes from a direct comparison with the binomial multitype branching process described above, and thus follows from Proposition 2.2 below, which bounds the total size of a multitype branching process. Consider a multitype branching process with  $d$  types, and let  $\mathbf{Z}_0, \mathbf{Z}_1, \mathbf{Z}_2, \dots$  be the generations of this process, where  $\mathbf{Z}_t$  is a vector whose  $i^{\text{th}}$  component,  $Z_t^i$ , is the number of type  $i$  individuals in the  $t^{\text{th}}$  generation.  $\mathbf{Z}_0$  is assumed to be deterministic, and all birth events are independent of one another. Let  $\mathbf{M} = (m_{ij})$  denote the matrix of expectations:

$$m_{ij} = \mathbb{E} \left( Z_1^j \mid \mathbf{Z}_0 = \mathbf{e}_i \right).$$

We say that a matrix,  $\mathbf{A}$ , or a vector,  $\mathbf{x}$ , is *positive* if  $A_{ij} > 0$  for all  $i$  and  $j$  or  $x_i > 0$  for all  $i$ . Lemma 2.1 is a simplified restatement of the Perron-Frobenius Theorem as it appears in [2].

**Lemma 2.1.** *If  $\mathbf{M}^N$  is positive for some natural number  $N$  then  $\mathbf{M}$  has a positive simple eigenvalue,  $\rho$ , that is greater in absolute value than any other eigenvalue, and  $\rho$  corresponds to a positive right eigenvector  $\boldsymbol{\mu} : \mathbf{M}\boldsymbol{\mu} = \rho\boldsymbol{\mu}$ .*

We wish to state Proposition 2.2 in some generality, but we will only be applying it to the multitype branching process with expectation matrix  $\mathbf{M}_\lambda$ . When  $d \geq 3$ , it is easy to check that  $\mathbf{M}_\lambda^2$  is positive, and when  $d = 2$  it can be verified directly that  $\mathbf{M}_\lambda$  has a unique positive eigenvalue corresponding to a positive right eigenvector.

**Proposition 2.2.** *If  $\mathbf{M}\boldsymbol{\mu} = \rho\boldsymbol{\mu}$  where  $0 < \rho < 1$  and  $\boldsymbol{\mu}$  is positive, and there exists  $\theta_0 > 0$  such that  $\mathbb{E} [e^{\theta_0 \langle \mathbf{Z}_1, \boldsymbol{\mu} \rangle} | \mathbf{Z}_0 = \mathbf{e}_i] < \infty$  for  $i = 1, \dots, d$ , then  $\mathbb{P} (\sum_{t=0}^{\infty} \|\mathbf{Z}_t\|_1 > x) \leq Ce^{-\alpha x}$  where  $\alpha, C > 0$  can be chosen as follows:*

$$e^{-\alpha} := \min_{\theta \in [0, \theta_0]} \max_i \mathbb{E} \left[ e^{\theta \langle \mathbf{Z}_1 - \mathbf{e}_i, \boldsymbol{\mu} \rangle} | \mathbf{Z}_0 = \mathbf{e}_i \right],$$

and, letting  $\theta'$  be the value of  $\theta$  for which this minimum is attained,  $C := e^{\theta' \langle \mathbf{Z}_0, \boldsymbol{\mu} \rangle}$ .

*Proof of Proposition 2.2.* Suppose  $X \geq 0$  is an arbitrary random variable such that  $\mathbb{E} e^{\theta_1 X} < \infty$  for some  $\theta_0 > 0$ . By expanding, for  $\theta < \theta_0$ :

$$\frac{d}{d\theta} (\mathbb{E} e^{\theta X}) \Big|_{\theta=0} = \frac{d}{d\theta} \left( 1 + \theta \mathbb{E} X + \frac{1}{2} \theta^2 \mathbb{E} X^2 + \dots \right) \Big|_{\theta=0} = \mathbb{E} X.$$

Thus, for each  $i = 1, \dots, d$ ,

$$\frac{d}{d\theta} \left( \mathbb{E} \left[ e^{\theta \langle \mathbf{Z}_1, \boldsymbol{\mu} \rangle} | \mathbf{Z}_0 = \mathbf{e}_i \right] \right) \Big|_{\theta=0} = \mathbb{E} [\langle \mathbf{Z}_1, \boldsymbol{\mu} \rangle | \mathbf{Z}_0 = \mathbf{e}_i] = \rho \mu_i. \quad (1)$$

Now consider a random walk version of the multitype branching process,  $(\mathbf{S}_t)$ , constructed as follows.  $\mathbf{S}_0 = \mathbf{Z}_0$  is a non-random initial vector whose  $i^{\text{th}}$  component indicates the number of type  $i$  individuals who are *active*. At each step, an active individual is chosen uniformly at random, it gives birth to a random number of individuals depending on its type and according to the law for that type in the branching process (these new individuals are considered active), then it is made inactive (thus no longer included in  $\mathbf{S}_t$ ).

$$\mathbf{S}_{t+1} = \mathbf{S}_t + \sum_{i=1}^d \mathbb{1}_{\{j_{t+1}=i\}} (\mathbf{X}_{t+1}^i - \mathbf{e}_i) \quad (2)$$

Where  $j_t$  is the random variable that takes the value  $i$  if an active individual of type  $i$  is selected at time  $t$ , so  $\mathbb{P}(j_{t+1} = i | \mathbf{S}_t = \mathbf{v}) = v_i / \|\mathbf{v}\|_1$ , and for each  $i = 1, \dots, d$  the random vectors  $\mathbf{X}_t^i$  are independent and equal in distribution to  $\mathbf{Z}_1$  conditional on  $\mathbf{Z}_0 = \mathbf{e}_i$ . Notice that  $j_{t+1}$  is dependent on  $\mathbf{S}_t$ , but  $\mathbf{X}_{t+1}^i$  is not. The process continues until the stopping time  $T := \inf \{t \mid \mathbf{S}_t = \mathbf{0}\}$ , at which time the process dies out. Notice that  $T = \sum_{t=0}^{\infty} \|\mathbf{Z}_t\|_1$ .

Now consider an increment of this process:

$$\begin{aligned}
\phi_{\mathbf{v}}(\theta) &:= \mathbb{E} \left[ \exp \left( \theta \left\langle \sum_i \mathbb{1}_{\{j_{t+1}=i\}} (\mathbf{X}_{t+1}^i - \mathbf{e}_i), \boldsymbol{\mu} \right\rangle \right) \middle| \mathbf{S}_t = \mathbf{v} \right] \\
&= \sum_i \frac{v_i}{\|\mathbf{v}\|_1} \mathbb{E} e^{\theta \langle \mathbf{X}_{t+1}^i - \mathbf{e}_i, \boldsymbol{\mu} \rangle} \\
&\leq \max_i \mathbb{E} e^{\theta \langle \mathbf{X}_1^i - \mathbf{e}_i, \boldsymbol{\mu} \rangle} \\
&=: \psi(\theta).
\end{aligned} \tag{3}$$

Note that the last expression does not depend on the previous state vector,  $\mathbf{v}$ . By equation (1), for each  $i = 1, \dots, d$  we have:

$$\frac{d}{d\theta} \left( \mathbb{E} e^{\theta \langle \mathbf{X}_1^i - \mathbf{e}_i, \boldsymbol{\mu} \rangle} \right) \Big|_{\theta=0} = \mu_i(\rho - 1).$$

Since  $\rho < 1$  and  $\mu_i > 0$  for all  $i$ , we have that there exist  $\theta', \alpha > 0$  such that  $\psi(\theta') = e^{-\alpha} < 1$ . The following computation shows that  $e^{\theta \langle \mathbf{S}_t, \boldsymbol{\mu} \rangle} / \psi(\theta)^t$  is a supermartingale, and uses the estimate in (3).

$$\begin{aligned}
\mathbb{E} \left[ e^{\theta \langle \mathbf{S}_{t+1}, \boldsymbol{\mu} \rangle} \middle| \mathbf{S}_t \right] &= \sum_{\mathbf{v} \geq \mathbf{0}} \mathbb{1}_{\{\mathbf{S}_t = \mathbf{v}\}} e^{\theta \langle \mathbf{v}, \boldsymbol{\mu} \rangle} \phi_{\mathbf{v}}(\theta) \\
&\leq \psi(\theta) e^{\theta \langle \mathbf{S}_t, \boldsymbol{\mu} \rangle}
\end{aligned}$$

So, by the Optional Stopping Theorem for positive supermartingales,

$$C := e^{\theta' \langle \mathbf{S}_0, \boldsymbol{\mu} \rangle} \geq \mathbb{E} \left[ \frac{e^{\theta' \langle \mathbf{S}_T, \boldsymbol{\mu} \rangle}}{\psi(\theta')^T} \right] = \mathbb{E} e^{\alpha T}$$

then by Markov's inequality:

$$\begin{aligned}
\mathbb{P}(T > x) &= \mathbb{P}(e^{\alpha T} > e^{\alpha x}) \\
&\leq C e^{-\alpha x}.
\end{aligned} \quad \square$$

*Proof of Theorem 1.1.* Consider a fixed vertex,  $\mathbf{v}$ , in the Hamming Torus, and let  $C_{\mathbf{v}}$  be the cluster containing that vertex ( $C_{\mathbf{v}} = \emptyset$  if the vertex  $\mathbf{v}$  is unoccupied). We can reveal the vertices in  $C_{\mathbf{v}}$  using the following cluster-discovering algorithm:

1. Initialize the sets  $R_0 = \emptyset$ ,  $A_0 = \{\mathbf{v}\}$ ,  $U_0 = V \setminus A_0$ .
2. Choose a vertex,  $\mathbf{v}_t$ , from  $A_t$  in some measurable way (e.g., lexicographically or uniformly at random).
3.  $A_{t+1} = A_t \setminus \{\mathbf{v}_t\} \cup \{\mathbf{w} \in \mathcal{N}(\mathbf{v}_t) \cap U_t \mid \mathbf{w} \text{ is occupied}\}$
4.  $U_{t+1} = U_t \setminus \mathcal{N}(\mathbf{v}_t)$
5.  $R_{t+1} = R_t \cup \{\mathbf{v}_t\}$
6. If  $A_{t+1}$  is empty, then return  $C_{\mathbf{v}} = R_{t+1}$ , otherwise increment  $t$  and go to step 2.

In the above algorithm,  $\mathcal{N}(\mathbf{v}) = \{\mathbf{w} \in V \mid d(\mathbf{v}, \mathbf{w}) = 1\}$  is the neighborhood of vertex  $\mathbf{v}$ . To compare this process with a multitype branching process, we can say that at step 3, a neighbor of  $\mathbf{v}_t$ , say  $\mathbf{w}$ , is of type  $i$  if  $\mathbf{v}_t - \mathbf{w} = m\mathbf{e}_i$  for some integer  $m$ . In future iterations of the algorithm, a type  $i$  individual can only yield offspring of types  $j \in \{1, \dots, d\} \setminus \{i\}$ , since all of its neighbors in the  $\mathbf{e}_i$  direction will have been removed from  $U_{t+1}$ . Thus,  $|C_{\mathbf{v}}|$  is bounded above (in distribution) by the size of a multitype branching process with  $Z_1^j$  distributed as a  $\text{Binomial}(a_j n, \frac{\lambda}{n})$  random variable for  $j \neq i$  and  $Z_1^i \equiv 0$  when  $\mathbf{Z}_0 = \mathbf{e}_i$ . The expectation matrix for this branching process is  $\mathbf{M}_\lambda$ , and the largest eigenvalue of  $\mathbf{M}_\lambda$  is less than 1 iff  $\lambda < \lambda_c$ . In this regime, we have that the moment generating functions for  $(\langle \mathbf{Z}_1 - \mathbf{e}_i, \boldsymbol{\mu} \rangle \mid \mathbf{Z}_0 = \mathbf{e}_i)$  are uniformly (in  $n$ ) bounded above for  $\theta \geq 0$ :

$$\begin{aligned} \mathbb{E} \left[ e^{\theta \langle \mathbf{Z}_1 - \mathbf{e}_i, \boldsymbol{\mu} \rangle} \mid \mathbf{Z}_0 = \mathbf{e}_i \right] &= e^{-\theta \mu_i} \prod_{j \neq i} \sum_{k_j=0}^{a_j n} \binom{a_j n}{k_j} \left( \frac{\lambda}{n} \right)^{k_j} \left( 1 - \frac{\lambda}{n} \right)^{a_j n - k_j} e^{\theta k_j \mu_j} \\ &= e^{-\theta \mu_i} \prod_{j \neq i} \left[ 1 + \frac{\lambda}{n} (e^{\theta \mu_j} - 1) \right]^{a_j n} \\ &\leq e^{-\theta \mu_i} \prod_{j \neq i} \exp [\lambda a_j (e^{\theta \mu_j} - 1)] \\ &=: \psi_i(\theta). \end{aligned}$$

It is easy to verify that  $\psi_i'(0) = \mu_i(\rho - 1) < 0$ ,  $\psi_i(0) = 1$ ,  $\psi_i(\theta) \rightarrow \infty$  as  $\theta \rightarrow \infty$ , and  $\psi_i''(\theta) > 0$  for all  $\theta > 0$  which imply that we can choose  $\alpha, C > 0$  such that

$$\begin{aligned} e^{-\alpha} &:= \min_{\theta \geq 0} \max_i \psi_i(\theta) =: \max_i \psi_i(\theta'), \\ C &:= \exp [\theta' \|\boldsymbol{\mu}\|_1]. \end{aligned}$$

Now applying Proposition 2.2 to this branching process yields the following inequality:

$$\begin{aligned} \mathbb{P} \left( |C_{\mathbf{v}}| > \frac{d+1}{\alpha} \log n \right) &\leq \mathbb{P} \left( \sum_{t=0}^{\infty} \|\mathbf{Z}_t\|_1 > \frac{d+1}{\alpha} \log n \right) \\ &\leq C n^{-d-1}. \end{aligned}$$

This implies the desired result:

$$\mathbb{P} \left( \max_{\mathbf{v} \in V} |C_{\mathbf{v}}| > \frac{d+1}{\alpha} \log n \right) \leq C(a_1 \cdots a_d) n^{-1}. \quad \square$$

### 3 Supercritical behavior

Theorem 1.2 shows that  $\lambda_c$  is the critical threshold for the emergence of a giant component. If we let  $q_i = \mathbb{P}(\mathbf{Z}_t = \mathbf{0} \text{ for some } t \mid \mathbf{Z}_0 = \mathbf{e}_i)$  be the extinction probabilities for the multitype branching process in which an individual of type  $i$  gives birth to a  $\text{Poisson}(\lambda a_j)$  number of type  $j$  individuals for  $j \neq i$  and zero individuals of type  $i$ , then the proportion of occupied vertices in the largest component when  $\lambda > \lambda_c$  is  $1 - (\prod_i q_i)^{1/(d-1)} \equiv 1 - q > 0$ .

**Theorem (1.2 restated).** Fix  $d \geq 2$  and  $a_1 \geq a_2 \geq \dots \geq a_d > 0$ , and let  $\lambda_c$  be the unique positive solution to  $\det(\mathbf{M}_{\lambda_c} - \mathbf{I}) = 0$ . If  $\lambda > \lambda_c$  and  $p = \frac{\lambda}{n}$  then the size of the largest connected component of the random site subgraph of the Hamming Torus is a.s.  $(1 - q)\lambda(\prod_i a_i) n^{d-1} + o(n^{d-1})$ . Furthermore, the second largest component has a.s.  $O(\log n)$  vertices.

Before beginning the proof, it will be useful to show that the matrix equation defining  $\lambda_c$  is equivalent to a polynomial equation that is easier to understand.

**Lemma (1.3 restated).**

$$\det(\mathbf{M}_\lambda - \mathbf{I}) = (-1)^d \left[ 1 - \sum_{\ell=2}^d (\ell-1)\lambda^\ell \sum_{\substack{S \subset \{1, \dots, d\} \\ |S|=\ell}} \prod_{i \in S} a_i \right]. \quad (4)$$

From this reformulation, it is easy to see that a positive solution to  $\det(\mathbf{M}_\lambda - \mathbf{I}) = 0$  exists and is unique, since the polynomial in the above expression is monotone increasing for odd  $d$  (monotone decreasing for even  $d$ ) and  $\lambda > 0$ , and the value at  $\lambda = 0$  is  $(-1)^d$ .

*Proof of Lemma 1.3.* To demonstrate the validity of equation (4), we will first make a change of variables by letting  $\gamma = 1/\lambda$ , then multiplying both sides by  $\gamma^d$ , so the equation we must verify is:

$$\begin{vmatrix} -\gamma & a_2 & a_3 & \cdots & a_d \\ a_1 & -\gamma & a_3 & \cdots & a_d \\ a_1 & a_2 & -\gamma & & \vdots \\ \vdots & \vdots & & \ddots & a_d \\ a_1 & a_2 & \cdots & a_{d-1} & -\gamma \end{vmatrix} = (-1)^d \left[ \gamma^d - \sum_{\ell=2}^d (\ell-1)\gamma^{d-\ell} \sum_{\substack{S \subset \{1, \dots, d\} \\ |S|=\ell}} \prod_{i \in S} a_i \right]. \quad (5)$$

We will proceed by induction on  $d$ . For  $d = 1$  both sides of equation (5) are equal to  $-\gamma$ . For  $d > 1$ , we will start by showing that the derivative with respect to  $\gamma$  on both sides of equation (5) is the same. Taking the derivative of the left hand side yields:

$$\begin{aligned}
\frac{d}{d\gamma} \begin{vmatrix} -\gamma & a_2 & a_3 & \cdots & a_d \\ a_1 & -\gamma & a_3 & \cdots & a_d \\ a_1 & a_2 & -\gamma & & \vdots \\ \vdots & \vdots & & \ddots & a_d \\ a_1 & a_2 & \cdots & a_{d-1} & -\gamma \end{vmatrix} &= \begin{vmatrix} -1 & 0 & 0 & \cdots & 0 \\ a_1 & -\gamma & a_3 & \cdots & a_d \\ a_1 & a_2 & -\gamma & & \vdots \\ \vdots & \vdots & & \ddots & a_d \\ a_1 & a_2 & \cdots & a_{d-1} & -\gamma \end{vmatrix} + \begin{vmatrix} -\gamma & a_2 & a_3 & \cdots & a_d \\ 0 & -1 & 0 & \cdots & 0 \\ a_1 & a_2 & -\gamma & & \vdots \\ \vdots & \vdots & & \ddots & a_d \\ a_1 & a_2 & \cdots & a_{d-1} & -\gamma \end{vmatrix} \\
&+ \cdots + \begin{vmatrix} -\gamma & a_2 & a_3 & \cdots & a_d \\ a_1 & -\gamma & a_3 & \cdots & a_d \\ \vdots & \vdots & \ddots & & \vdots \\ a_1 & a_2 & \cdots & -\gamma & a_d \\ 0 & 0 & \cdots & 0 & -1 \end{vmatrix} \\
&= - \sum_{j=1}^d (-1)^{d-1} \left[ \gamma^{d-1} - \sum_{\ell=2}^{d-1} (\ell-1) \gamma^{d-1-\ell} \sum_{\substack{S \subset \{1, \dots, d\} \setminus \{j\} \\ |S|=\ell}} \prod_{i \in S} a_i \right] \\
&= (-1)^d \left[ d\gamma^{d-1} - \sum_{\ell=2}^{d-1} (\ell-1) \gamma^{d-1-\ell} \sum_{j=1}^d \sum_{\substack{S \subset \{1, \dots, d\} \setminus \{j\} \\ |S|=\ell}} \prod_{i \in S} a_i \right] \\
&= (-1)^d \left[ d\gamma^{d-1} - \sum_{\ell=2}^{d-1} (\ell-1) \gamma^{d-1-\ell} (d-\ell) \sum_{\substack{S \subset \{1, \dots, d\} \\ |S|=\ell}} \prod_{i \in S} a_i \right]
\end{aligned}$$

which is the derivative of the right hand side of equation (5). The second line follows from the inductive hypothesis. The last line results from counting the number of values of  $j$  such that, given a fixed set  $S \subset \{1, \dots, d\}$  of size  $\ell$ ,  $S$  is also a subset of  $\{1, \dots, d\} \setminus \{j\}$  of size  $\ell$ . This is the same as counting the number of values of  $j$  such that  $j \notin S$ , which is  $(d-\ell)$ . Now that we have shown the derivatives to be equivalent, to verify equation (5) we only need to show that it holds for  $\gamma = 0$ . To this end, we can factor the matrix on the left hand side of (5) when  $\gamma = 0$  as:

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \\ -1 & -2 & \cdots & -(d-1) & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \\ 0 & \cdots & 0 & 0 & (d-1) \end{pmatrix} \\
\times \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_d \end{pmatrix}.$$

Since all of these matrices are triangular, it is clear that the determinant on the left hand side of (5) when  $\gamma = 0$  is  $(-1)^{d-1}(d-1) \prod_i a_i$ , which is equal to the right hand side of (5) when  $\gamma = 0$ . Thus, we have demonstrated equation (4).  $\square$

*Proof of Theorem 1.2.* We begin by considering  $C_v$ , the cluster containing a fixed vertex  $v$ , and reveal the vertices in  $C_v$  using a similar algorithm to that described in the proof of Theorem 1, but with one noteworthy modification: We will remove some additional vertices at step 4 before they can be observed in any subsequent iterations. This will avoid the problem of generating closed loops, which would severely reduce the unseen neighborhoods of active vertices and make it difficult to couple the process with a lower-bounding branching process. The cluster-discovering algorithm is then:

1. Initialize the sets  $R_0 = \emptyset$ ,  $A_0 = \{v\}$ ,  $U_0 = V \setminus A_0$ .
2. Choose a vertex,  $v_t$ , from  $A_t$  in some measurable way (e.g., lexicographically or uniformly at random).
3.  $A_{t+1} = A_t \setminus \{v_t\} \cup \{w \in \mathcal{N}(v_t) \cap U_t \mid w \text{ is occupied}\}$
4.  $U_{t+1} = (U_t \setminus \mathcal{N}(v_t)) \setminus \{w \in \mathcal{N}(u) \cap \mathcal{N}(v) \mid u, v \in A_{t+1}, u \neq v\}$
5.  $R_{t+1} = R_t \cup \{v_t\}$
6. If  $A_{t+1}$  is empty, then return  $R_{t+1}$ , otherwise increment  $t$  and go to step 2.

We will need to keep track of the number of vertices of each type, so let  $\mathcal{A}_t^{(i)}$  denote the number of type  $i$  vertices in  $A_t$ , let  $\mathbf{A}_t = (\mathcal{A}_t^{(1)}, \dots, \mathcal{A}_t^{(d)})$ , and let  $\mathcal{N}^i(v) = \{w \in V \mid v - w = m\mathbf{e}_i \text{ for some } m\}$ . As explained in the proof of Theorem 1.1, a vertex,  $w$ , discovered at time  $t$  is labeled type  $i$  if  $w \in \mathcal{N}^i(v_t)$ . That is, a vertex is of type  $i$  if it is discovered by searching neighboring vertices in the direction of the  $i^{\text{th}}$  basis vector.

The above algorithm will give a lower bound on  $C_v$ , but we will make use of an upper-bounding branching process, as in the proof of Theorem 1.1, to tightly control the size of  $C_v$ . We can couple the random walk version of the upper-bounding branching process with  $\mathbf{A}_t$  by using the same random variables in the construction of  $\mathbf{S}_t$  as in the construction of  $\mathbf{A}_t$  whenever possible, and when independence is an issue, we just add independent copies of these random variables to  $\mathbf{S}_t$ . More rigorously, let  $\xi_v$  be the random variable that takes the value 1 if the vertex  $v$  is occupied and is 0 otherwise. Then we can define  $\mathbf{S}_t$  iteratively for each  $i$ :

$$S_{t+1}^{(i)} = S_t^{(i)} + \sum_{w \in \mathcal{N}^i(v_t) \cap U_t} \xi_w + \sum_{k=1}^{a_i n - |\mathcal{N}^i(v_t) \cap U_t|} \eta_k^{(t,i)} - \mathbb{1}_{\{v_t \text{ is of type } i\}}, \quad (6)$$

where the  $\eta_k^{(t,i)}$  are i.i.d. Bernoulli( $p$ ) random variables, and are independent of the  $\xi_w$ . As long as  $\mathbf{A}_t \neq \mathbf{0}$ , then we can choose the same focal vertex  $v_t$  for both processes. But, if  $\mathbf{A}_t = \mathbf{0}$  and  $\mathbf{S}_t \neq \mathbf{0}$  then we can just choose a vertex from the process  $\mathbf{S}_t$  in some measurable way.

### 3.1 Existence and Survival of a Lower-bounding Random Walk

We will make use of this upper bounding random walk in a moment, but first we also need to introduce a lower bounding random walk - also based on a branching process. We will use the upper

bounding random walk to grow a sufficiently large cluster for the lower bounding walk to take over, but first we will need to know what “sufficiently large” means. Also, we need to know that such a lower bound exists, so we begin there. Ideally, we would like to have a lower bound on  $|U_r \cap \mathcal{N}^i(\mathbf{v})|$  for some large enough  $r$ . Using the Erdős-Rényi model as guidance, we need  $n^{(d-1)/2} \ll r \ll n^{d-1}$ , so we choose  $r = n^{d-\frac{4}{3}}$ .

To generate the lower bounding random walk, we will begin by letting  $p = p_1 + p_2 - p_1 p_2$ , where  $p_1 = \frac{\lambda_1}{n}$  and  $p = \frac{\lambda}{n}$ . We want to choose  $\lambda_1 < \lambda$  so that  $\lambda_1 > \lambda_c$ , but  $\lambda_1$  is smaller than the positive solution to  $\det(\mathbf{M}_\lambda^{(d-1)} - \mathbf{I}^{(d-1)}) = 0$ , where  $\mathbf{M}_\lambda^{(d-1)}$  denotes the submatrix of  $\mathbf{M}_\lambda$  consisting of the first  $(d-1)$  rows and columns. Call this solution  $\lambda_c^{(d-1)}$ . In other words, we want  $\lambda_1$  to be supercritical for the  $d$ -dimensional process, but subcritical for the process restricted to any  $(d-1)$ -dimensional plane. By the monotonicity of the process of growing a cluster, the critical value for the  $(d-1)$ -dimensional process obtained by considering the first  $(d-1)$  dimensions is smallest, since we assumed that  $a_1 \geq a_2 \geq \dots \geq a_d > 0$ . The critical value for this process is  $\lambda_c^{(d-1)}$ , so to guarantee the existence of such a  $\lambda_1$ , we just need to verify that  $\lambda_c < \lambda_c^{(d-1)}$ . This is easiest to see by using equation (4). The values of  $\det(\mathbf{M}_{\lambda_c^{(d-1)}} - \mathbf{I})$  and  $\det(\mathbf{M}_0 - \mathbf{I})$  have opposite signs, so by the Intermediate Value Theorem and the uniqueness of  $\lambda_c$ , we have that  $0 < \lambda_c < \lambda_c^{(d-1)}$ .

Consider the process of discovering a cluster starting with  $m = O(\log n)$  vertices (we will decide what  $m$  is exactly just before equation (13)) of arbitrary types in which each unseen vertex has probability  $p_1$  of being occupied; call this process  $\mathcal{A}'_t$ . The corresponding active set of vertices will be referred to as  $A'_t$ , and we have assumed that  $|A'_0| = m$ . These  $m$  vertices will ultimately come from running the cluster-discovering process up to some time  $s = \beta \log n$  starting from a single vertex and using parameter  $p$  in Section 3.2. Conditioned on the survival of this process, we will have that  $|R'_0| = |R_s| = s$ , where  $R'_t$  is the set of occupied vertices which have been removed from the active set up to time  $t$  when we start with  $m$  active vertices and use parameter  $p_1$ . Also, let  $U'_t$  be the set of unobserved vertices up to time  $t$ , and  $U'_0 = U_s$ . Denote by

$$\mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)} := \{\mathbf{v} \in V \mid v_{i_1} = \ell_1, \dots, v_{i_k} = \ell_k\} \quad (7)$$

the intersection of the vertex set,  $V$ , with the planes  $v_{i_1} = \ell_1, \dots, v_{i_k} = \ell_k$ . Throughout this argument we will assume that  $1 \leq i_1, \dots, i_k \leq d$ ,  $i_{j_1} \neq i_{j_2}$  for  $j_1 \neq j_2$ , and  $\ell_j \in \{1, \dots, a_{i_j} n\}$ , so we have no trivial constraints. We want to bound the number of vertices that become occupied up to time  $r$  in any 2-dimensional plane,  $\mathcal{P}_{(i_1, \dots, i_{d-2})}^{(\ell_1, \dots, \ell_{d-2})}$ . This will give us the desired lower bound on  $|U'_r \cap \mathcal{N}^i(\mathbf{v}')|$  for each  $i$  and  $\mathbf{v}'$  that appear in the cluster discovering algorithm. To get a bound on the number of occupied vertices in these 2-dimensional planes, we will work backwards by first getting a bound on the number of vertices that become occupied up to time  $r$  in  $(d-1)$ -dimensional planes of the form  $\mathcal{P}_{(i_1)}^{(\ell_1)}$ . In the statement of Lemma 3.1, we allow  $m$  and  $s$  to be larger than we will need for the proof of Theorem 1.2. This is because we will also use Lemma 3.1 in the proofs of Theorems 1.5 and 1.6, where  $m$  and  $s$  will be slightly larger.

**Lemma 3.1.** *If  $|A'_0| = m = o(n^{2/3})$ ,  $|R'_0| = s = o(n^{2/3})$  and  $r = n^{d-4/3}$ , then there is a constant  $K_d$  such that:*

$$\begin{aligned} \mathbb{P}_{p_1} \left( \left| \mathcal{P}_{(i_1, \dots, i_{d-2})}^{(\ell_1, \dots, \ell_{d-2})} \cap (A'_r \cup R'_r) \right| \geq K_d n^{2/3} (\log n)^{d-2} \text{ for some } (i_1, \dots, i_{d-2}) \text{ and } (\ell_1, \dots, \ell_{d-2}) \right) \\ = O(n^{-(d+1)}). \end{aligned} \quad (8)$$

*Proof of Lemma 3.1.* We begin with  $d = 2$ . In this case, equation (8) is simply the statement that at most  $K_2 n^{2/3}$  occupied vertices have been discovered by time  $r = n^{2/3}$ , so we merely need a large

deviation bound. At each step, we can discover at most a  $\text{Binomial}(a_1 n, \frac{\lambda_1}{n})$  number of occupied vertices. Thus,

$$\begin{aligned} \mathbb{P}\left(|A'_r \cup R'_r| \geq 2a_1 \lambda_1 n^{2/3}\right) &\leq e^{-2a_1 \lambda_1 r} \mathbb{E} e^{|A'_r \cup R'_r|} \\ &\leq e^{-2a_1 \lambda_1 r + m + s} \left[1 + \frac{\lambda_1}{n}(e-1)\right]^{a_1 n r} \\ &\leq \exp[-2a_1 \lambda_1 r + a_1 \lambda_1 (e-1)r + m + s] \\ &= O(n^{-3}). \end{aligned}$$

In the first line above we exponentiated and applied Markov's inequality, in the second line we used that the total number of births up to time  $r$  is stochastically bounded by a  $\text{Binomial}(a_1 n r, \frac{\lambda_1}{n})$  random variable, in the fourth line we used the bound  $1 + x \leq e^x$ , and in the last line we used that  $m + s = o(n^{2/3})$  and  $e - 1 < 2$ . Thus we have (8) for  $d = 2$  with  $K_2 = 2a_1 \lambda_1$ .

Now, for  $d \geq 3$ , first consider a fixed plane  $\mathcal{P}_{(i_1)}^{(\ell_1)}$ . At each step in the cluster discovering algorithm, either a vertex in this plane is chosen or a vertex in some other parallel plane is chosen. If a vertex in this plane is chosen, then if we restrict our attention to just this plane, that is we remove  $\mathcal{N}^{i_1}(\mathbf{v}'_t)$  from  $\mathcal{N}(\mathbf{v}'_t)$  at step 3 of the algorithm, the resulting process will look like a subcritical branching process in  $(d-1)$  dimensions, by the assumptions we made on  $\lambda_1$ . Thus, restricted only to this plane, no vertex will give rise to a cluster larger than  $\beta_1 \log n$  vertices with at least probability  $(1 - n^{-d-2})$  (take  $\beta_1 = \frac{2d+1}{\alpha_{(d-1)}}$  in the proof of Theorem 1). If, instead, a vertex in another plane is chosen, then this vertex may have a single opportunity to give rise to an occupied "seed" vertex in  $\mathcal{P}_{(i_1)}^{(\ell_1)}$ . If the potential seed vertex has already been examined during a previous iteration of the algorithm, then it will not increase the number of seeds in  $\mathcal{P}_{(i_1)}^{(\ell_1)}$  whether it is occupied or not. If the potential seed vertex has not been examined (it is in  $U_t$ ), then it will be occupied with probability  $\frac{\lambda_1}{n}$ . So by time  $r$ , the expected number of seeds is  $\leq \frac{\lambda_1 r}{n}$ , and each seed is associated with a cluster restricted to  $\mathcal{P}_{(i_1)}^{(\ell_1)}$  of size at most  $\beta_1 \log n$ . If we let  $X_r$  be the number of seed vertices generated in  $\mathcal{P}_{(i_1)}^{(\ell_1)}$  by time  $r$ , then  $X_r$  is a sum of a random number of Bernoulli( $\frac{\lambda_1}{n}$ ) random variables. But  $X_r$  is stochastically bounded above by  $X'_r \sim \text{Binomial}(r, \frac{\lambda_1}{n})$ , since the number of potential seed vertices cannot exceed  $r$ . Using this fact, and applying a standard generating function argument shows that for  $\gamma_1 = 2 \log 2 - 1 > 0$ ,

$$\mathbb{P}\left(X_r \geq 2 \frac{\lambda_1 r}{n}\right) \leq e^{-\gamma_1 \frac{\lambda_1 r}{n}}.$$

Recalling that we have chosen  $r = n^{d-\frac{4}{3}}$ , this implies that

$$\mathbb{P}\left(\left|\mathcal{P}_{(i_1)}^{(\ell_1)} \cap (A'_r \cup R'_r)\right| \geq 2\lambda_1 \beta_1 n^{d-7/3} \log n\right) \leq e^{-\gamma_1 \lambda_1 n^{d-(7/3)}} + n^{-d-2}$$

and since the number of planes,  $\mathcal{P}_{(i_1)}^{(\ell_1)}$ , is  $\leq da_1 n$ , that

$$\mathbb{P}\left(\left|\mathcal{P}_{(i_1)}^{(\ell_1)} \cap (A'_r \cup R'_r)\right| \geq 2\lambda_1 \beta_1 n^{d-7/3} \log n \text{ for some } (i_1) \text{ or } (\ell_1)\right) = O(n^{-(d+1)}). \quad (9)$$

Now we wish to show that if every plane of the form  $\mathcal{P}_{(i_1, \dots, i_{k-1})}^{(\ell_1, \dots, \ell_{k-1})}$  has at most  $N$  occupied vertices by time  $r$ , then with high probability, every plane of the form  $\mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)}$  has at most  $2k\lambda_1 \frac{N}{n} \beta_k \log n$  occupied vertices by time  $r$ . The reasoning is similar to the argument above for  $\mathcal{P}_{(i_1)}^{(\ell_1)}$ . Since no

plane of the form  $\mathcal{P}_{(i_1, \dots, i_{k-1})}^{(\ell_1, \dots, \ell_{k-1})}$  has more than  $N$  occupied vertices, then certainly no set of the form  $\mathcal{P}_{(i_1, \dots, i_{k-1})}^{(\ell_1, \dots, \ell_{k-1})} \setminus \mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)}$  has more than  $N$  occupied vertices. Let us fix  $\mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)}$ , and denote by

$$\mathcal{P}_{(i_1, \dots, i_k) \setminus (i_j)}^{(\ell_1, \dots, \ell_k) \setminus (\ell_j)} := \mathcal{P}_{(i_1, \dots, i_{j-1}, i_{j+1}, \dots, i_k)}^{(\ell_1, \dots, \ell_{j-1}, \ell_{j+1}, \dots, \ell_k)} \setminus \mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)}$$

In one iteration of the cluster discovering algorithm, the focal vertex,  $\mathbf{v}'_t$ , is either in  $\mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)}$ , in  $\mathcal{P}_{(i_1, \dots, i_k) \setminus (i_j)}^{(\ell_1, \dots, \ell_k) \setminus (\ell_j)}$  for some  $j = 1, \dots, k$ , or neither. If it is in neither of these sets, then  $\mathcal{N}(\mathbf{v}_t) \cap \mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)} = \emptyset$ , so this case does not contribute to the number of occupied vertices in this plane. If the focal vertex is in  $\mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)}$ , then we can restrict our attention to just this plane, and the process looks like a subcritical process in  $d - k$  dimensions (by our assumptions on  $\lambda_1$ ). That is, we remove  $\cup_{j=1}^k \mathcal{N}^{i_j}(\mathbf{v}'_t)$  from  $\mathcal{N}(\mathbf{v}'_t)$  at Step 3 of the cluster discovering algorithm. Therefore, all vertices in  $\mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)}$  belong to clusters restricted to  $\mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)}$  of size at most  $\beta_k \log n$  with probability at least  $(1 - n^{-(d+k+1)})$  (take  $\beta_k = \frac{2d+1}{\alpha_{(d-k)}}$ ). Finally, if the focal vertex is in  $\mathcal{P}_{(i_1, \dots, i_k) \setminus (i_j)}^{(\ell_1, \dots, \ell_k) \setminus (\ell_j)}$  for some  $j = 1, \dots, k$ , then it has precisely one neighbor in  $\mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)}$ . If this neighbor is unseen (is in  $U_t$ ), then it has probability  $\frac{\lambda_1}{n}$  of being an occupied seed vertex. There are at most  $kN$  occupied vertices in  $\mathcal{P}_{(i_1, \dots, i_k) \setminus (i_j)}^{(\ell_1, \dots, \ell_k) \setminus (\ell_j)}$  for  $j = 1, \dots, k$ , so the number of seed vertices,  $Y_N^k$ , is a sum of at most  $kN$  Bernoulli( $\frac{\lambda_1}{n}$ ) random variables, and is thus stochastically bounded above by a Binomial( $kN, \frac{\lambda_1}{n}$ ) random variable. Thus, we find that

$$\mathbb{P} \left( Y_N^k \geq 2 \frac{\lambda_1 k N}{n} \right) \leq e^{-\gamma_1 \frac{\lambda_1 k N}{n}},$$

where, again,  $\gamma_1 = 2 \log 2 - 1$ . So, conditional on the event that no plane of the form  $\mathcal{P}_{(i_1, \dots, i_{k-1})}^{(\ell_1, \dots, \ell_{k-1})}$  has more than  $N$  occupied vertices by time  $r$ , we have that:

$$\mathbb{P} \left( \mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)} \text{ has } \geq 2k\lambda_1\beta_k \frac{N}{n} \log n \text{ occupied vertices by time } r \right) \leq e^{-\gamma_1 \lambda_1 k \frac{N}{n}} + n^{-d-k-1}.$$

Since the number of planes,  $\mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)}$ , is at most  $\binom{d}{k} a_1^k n^k$ , and provided that  $N = \omega(n^{1+\epsilon})$  for some  $\epsilon > 0$ , we have:

$$\mathbb{P} \left( \text{Any } \mathcal{P}_{(i_1, \dots, i_k)}^{(\ell_1, \dots, \ell_k)} \text{ has } \geq 2k\lambda_1\beta_k \frac{N}{n} \log n \text{ occupied vertices by time } r \right) = O(n^{-(d+1)}) \quad (10)$$

conditional on the event that no plane of the form  $\mathcal{P}_{(i_1, \dots, i_{k-1})}^{(\ell_1, \dots, \ell_{k-1})}$  has more than  $N$  occupied vertices by time  $r$ .

Finally, we can put together equations (9) and (10) to get:

$$\mathbb{P} \left( \text{Any } \mathcal{P}_{(i_1, \dots, i_{d-2})}^{(\ell_1, \dots, \ell_{d-2})} \text{ has } \geq K_d n^{2/3} (\log n)^{d-2} \text{ occupied vertices by time } r \right) = O(n^{-(d+1)})$$

where  $K_d = 2^{d-2} \cdot (d-2)! \cdot \lambda_1^{d-2} \cdot \left( \prod_{k=1}^{d-2} \beta_k \right)$  □

So with probability at least  $1 - O(n^{-(d+1)})$ , for  $t \leq r$  and  $\mathbf{v}_t$  of type  $j \neq i$ ,

$$\begin{aligned} |U'_t \cap \mathcal{N}^i(\mathbf{v}'_t)| &\geq a_i n - K_d n^{2/3} (\log n)^{d-2} \\ &\geq a_i (1 - \delta) n \end{aligned}$$

for sufficiently large  $n$ , where  $\delta > 0$  is such that  $(1 - \delta)\lambda_1 > \lambda_c$ . This implies that, up to time  $r$ , our process of discovering occupied vertices is bounded below by the random walk version of a multitype branching process with expectation matrix  $\mathbf{M}_{(1-\delta)\lambda_1}$ . Let us call the random walk version of this multitype branching process  $\mathbf{W}_t = (W_t^{(1)}, \dots, W_t^{(d)})$ , and assume that  $\mathbf{W}_0 = \mathcal{A}'_0$ . We can couple the processes  $\mathbf{W}_t$  and  $\mathcal{A}'_t$  until time  $T_0 \wedge T_\delta \wedge r$ , where

$$\begin{aligned} T_0 &:= \inf \{t \mid \mathbf{W}_t = \mathbf{0}\}, \\ T_\delta &:= \inf \{t \mid |U_t \cap \mathcal{N}^i(\mathbf{v}'_t)| < a_i(1 - \delta)n \text{ for } \mathbf{v}'_t \text{ of type } j \text{ and some } i \neq j\}, \end{aligned}$$

by first choosing  $\mathbf{v}'_t$  as follows: first decide that  $\mathbf{v}'_t$  will be of type  $i$  with probability  $\frac{W_t^{(i)}}{\|\mathbf{W}_t\|_1}$ , then choose randomly a vertex from  $A'_t$  that is of this type. Let  $\mathcal{R}_\delta^{(i)}(\mathbf{v}'_t)$  be the unseen neighborhood of  $\mathbf{v}'_t$  restricted to the first (lexicographically)  $a_i(1 - \delta)n$  elements of  $U_t \cap \mathcal{N}^i(\mathbf{v}'_t)$ . Then we can complete the coupling of  $\mathbf{W}_t$  with  $\mathcal{A}'_t$  as

$$\mathcal{A}'_{t+1} = \mathcal{A}'_t + \sum_{\mathbf{w} \in U_t \cap \mathcal{N}^i(\mathbf{v}'_t)} \xi'_{\mathbf{w}} - \mathbb{1}_{\{\mathbf{v}'_t \text{ is of type } i\}} \quad (11)$$

$$W_{t+1}^{(i)} = W_t^{(i)} + \sum_{\mathbf{w} \in \mathcal{R}_\delta^{(i)}(\mathbf{v}'_t)} \xi'_{\mathbf{w}} - \mathbb{1}_{\{\mathbf{v}'_t \text{ is of type } i\}} \quad (12)$$

where  $\xi'_{\mathbf{w}}$  is the random variable that takes the value 1 if vertex  $\mathbf{w}$  is occupied in the random site subgraph obtained using the probability parameter  $p_1$ , and is 0 otherwise.

We wish to show now that this coupling will last until time  $r$  if  $m$  is large enough. This will occur whenever the branching process corresponding to  $\mathbf{W}_t$  survives. Let

$$q'_i = \mathbb{P}(\mathbf{W}_t = \mathbf{0} \text{ for some } t \mid \mathbf{W}_0 = \mathbf{e}_i)$$

for  $i = 1, \dots, d$ . Since this branching process is supercritical,  $q'_i < 1$  for each  $i$ . If we let  $e^{-\gamma_2} = \max_i q'_i$  for  $\gamma_2 > 0$ , then

$$\begin{aligned} \mathbb{P}(\mathbf{W}_t = \mathbf{0} \text{ for some } t \mid \|\mathbf{W}_0\|_1 = m) &\leq \left(\max_i q'_i\right)^m \\ &= e^{-\gamma_2 m}. \end{aligned}$$

So if  $m = \frac{d+1}{\gamma_2} \log n$ , then the coupling between  $\mathbf{W}_t$  and  $\mathcal{A}'_t$  will last until time  $r$  with at least probability  $1 - n^{-(d+1)}$ . Thus, conditional on  $|A'_0| \geq m$ ,  $|A'_0| = o(n^{2/3})$  and  $|R'_0| = o(n^{2/3})$  we have

$$\mathbb{P}_{p_1}(|A'_r| = 0) = O\left(n^{-(d+1)}\right). \quad (13)$$

### 3.2 Establishing a Set of $m$ Active Vertices

As in the proof of Theorem 1.1, we make use of an upper-bounding branching process, but this time in a more intimate fashion. We couple the random walk version of the upper-bounding branching process,  $\mathbf{S}_t$ , with  $\mathcal{A}_t$  as in equation (6). Recall from equation (2) that we can write

$$\mathbf{S}_{t+1} \stackrel{d}{=} \mathbf{S}_t + \sum_{i=1}^d \mathbb{1}_{\{j_{t+1}=i\}} (\mathbf{X}_{t+1}^i - \mathbf{e}_i),$$

where  $j_{t+1}$  is the random variable that takes the value  $i$  if  $\mathbf{v}_t$  is of type  $i$  (or, when  $\mathbf{A}_t = \mathbf{0}$ , if a type  $i$  individual is chosen at time  $t+1$  in the random walk process), and where the  $\mathbf{X}_t^i$  are independent and distributed as  $\mathbf{Z}_1$  conditional on  $\mathbf{Z}_0 = \mathbf{e}_i$ . For this process (and later for the lower bounding process,  $\mathbf{W}_t$ ) we will need the following large deviations bounds.

**Lemma 3.2.** *Let  $\mathbf{S}_t$  be the random walk version of a branching process with  $d$  types in which an individual of type  $i$  has a  $\text{Binomial}(a_j n, \frac{\lambda}{n})$  number of offspring of type  $j \neq i$  and zero offspring of type  $i$ . If  $\mathbf{M}_\lambda \boldsymbol{\mu} = \rho \boldsymbol{\mu}$ , where  $\rho > 1$  and  $\boldsymbol{\mu}$  is the corresponding positive eigenvector normalized so that  $\|\boldsymbol{\mu}\|_1 = 1$ , then for  $y < 1 < x$*

$$\mathbb{P}(\langle \mathbf{S}_t - \mathbf{S}_0, \boldsymbol{\mu} \rangle \geq x(\rho - 1)\mu_{\max} t) \leq e^{-\eta_1 t} \quad (14)$$

$$\mathbb{P}(\langle \mathbf{S}_t - \mathbf{S}_0, \boldsymbol{\mu} \rangle \leq y(\rho - 1)\mu_{\min} t) \leq e^{-\eta_2 t} \quad (15)$$

where  $\mu_{\max} := \max_i(\mu_i)$ ,  $\mu_{\min} := \min_i(\mu_i)$ , and  $\eta_1, \eta_2 > 0$  depend on  $x$  and  $y$ , respectively.

*Proof of Lemma 3.2.* We will first prove inequality (14). First, as in the proof of Proposition (2.2), we consider an increment of the process  $\langle \mathbf{S}_t - \mathbf{S}_0, \boldsymbol{\mu} \rangle$ :

$$\begin{aligned} \phi_{\mathbf{v}}(\theta) &:= \mathbb{E} \left[ \exp \left( \theta \left\langle \sum_i \mathbb{1}_{\{j_{t+1}=i\}} (\mathbf{X}_{t+1}^i - \mathbf{e}_i), \boldsymbol{\mu} \right\rangle \right) \middle| \mathbf{S}_t = \mathbf{v} \right] \\ &= \sum_i \mathbb{P}(j_{t+1} = i \mid \mathbf{S}_t = \mathbf{v}) \mathbb{E} e^{\theta \langle \mathbf{X}_{t+1}^i - \mathbf{e}_i, \boldsymbol{\mu} \rangle} \\ &\leq \max_i \mathbb{E} e^{\theta \langle \mathbf{X}_1^i - \mathbf{e}_i, \boldsymbol{\mu} \rangle} \\ &=: \psi(\theta). \end{aligned}$$

Using this estimate, which is independent of the previous state vector,  $\mathbf{v}$ , we see that

$$\begin{aligned} \mathbb{E} \left[ e^{\theta \langle \mathbf{S}_{t+1}, \boldsymbol{\mu} \rangle} \middle| \mathbf{S}_t \right] &= \sum_{\mathbf{v} \geq \mathbf{0}} \mathbb{1}_{\{\mathbf{S}_t = \mathbf{v}\}} e^{\theta \langle \mathbf{v}, \boldsymbol{\mu} \rangle} \phi_{\mathbf{v}}(\theta) \\ &\leq \psi(\theta) e^{\theta \langle \mathbf{S}_t, \boldsymbol{\mu} \rangle}. \end{aligned}$$

This implies that

$$\mathbb{E} e^{\theta \langle \mathbf{S}_t - \mathbf{S}_0, \boldsymbol{\mu} \rangle} \leq [\psi(\theta)]^t. \quad (16)$$

By equation (1), for each  $i = 1, \dots, d$

$$\begin{aligned} \frac{d}{d\theta} \left( \mathbb{E} e^{\theta \langle \mathbf{X}_1^i - \mathbf{e}_i, \boldsymbol{\mu} \rangle} \right) \Big|_{\theta=0} &= (\rho - 1)\mu_i \\ &< x(\rho - 1)\mu_{\max} \\ &= \frac{d}{d\theta} \left( e^{\theta x(\rho - 1)\mu_{\max}} \right) \Big|_{\theta=0} \end{aligned}$$

since  $\rho > 1$  and  $\boldsymbol{\mu} > \mathbf{0}$ . Because  $\psi(0) = 1$ , this implies that there exist  $\theta_1, \eta_1 > 0$  such that

$$e^{-\eta_1} := \psi(\theta_1) e^{-\theta_1 x(\rho - 1)\mu_{\max}} < 1.$$

Thus, by Markov's inequality and inequality (16):

$$\begin{aligned} \mathbb{P}(\langle \mathbf{S}_t - \mathbf{S}_0, \boldsymbol{\mu} \rangle \geq x(\rho - 1)\mu_{\max} t) &= \mathbb{P} \left( e^{\theta_1 \langle \mathbf{S}_t - \mathbf{S}_0, \boldsymbol{\mu} \rangle} \geq e^{\theta_1 x(\rho - 1)\mu_{\max} t} \right) \\ &\leq \mathbb{E} e^{\theta_1 \langle \mathbf{S}_t - \mathbf{S}_0, \boldsymbol{\mu} \rangle} e^{-\theta_1 x(\rho - 1)\mu_{\max} t} \\ &\leq e^{-\eta_1 t}. \end{aligned}$$

So we have proven inequality (14). To prove inequality (15), we need the following estimate for each  $i = 1, \dots, d$ :

$$\begin{aligned} \frac{d}{d\theta} \left( \mathbb{E} e^{-\theta \langle \mathbf{X}_1^i - \mathbf{e}_i, \boldsymbol{\mu} \rangle} \right) \Big|_{\theta=0} &= -(\rho - 1)\mu_i \\ &< -y(\rho - 1)\mu_{\min} \\ &= \frac{d}{d\theta} \left( e^{-\theta y(\rho-1)\mu_{\min}} \right) \Big|_{\theta=0} \end{aligned}$$

since  $\rho > 1$  and  $\boldsymbol{\mu} > \mathbf{0}$ . Because  $\psi(0) = 1$ , this implies that there exist  $\theta_2, \eta_2 > 0$  such that

$$e^{-\eta_2} := \psi(-\theta_2) e^{\theta_2 y(\rho-1)\mu_{\min}} < 1.$$

Thus, by Markov's inequality and inequality (16):

$$\begin{aligned} \mathbb{P}(\langle \mathbf{S}_t - \mathbf{S}_0, \boldsymbol{\mu} \rangle \leq y(\rho - 1)\mu_{\min} t) &= \mathbb{P}\left(e^{-\theta_2 \langle \mathbf{S}_t - \mathbf{S}_0, \boldsymbol{\mu} \rangle} \geq e^{-\theta_2 y(\rho-1)\mu_{\min} t}\right) \\ &\leq \mathbb{E} e^{-\theta_2 \langle \mathbf{S}_t - \mathbf{S}_0, \boldsymbol{\mu} \rangle} e^{\theta_2 y(\rho-1)\mu_{\min} t} \\ &\leq e^{-\eta_2 t}. \end{aligned} \quad \square$$

We are now ready to prove the following lemma:

**Lemma 3.3.** *There exists a constant  $\beta$  depending on  $a_1, \dots, a_d$  such that if  $s = \beta \log n$  then*

$$\mathbb{P}(0 < \|\mathbf{A}_s\|_1 \leq m) = O\left(n^{-(d+1)}\right).$$

and

$$\mathbb{P}(\|\mathbf{A}_s\|_1 \geq K \log n) = O\left(n^{-(d+1)}\right).$$

where  $K \log n > m$ .

*Proof of Lemma 3.3.* We wish to apply Lemma 3.2 to  $\mathbf{S}_t$ , but the first step of this process is not the same as the subsequent steps, so we must handle this case separately. Notice that in the first step of the cluster-discovering algorithm, the initial vertex does not have a type, and is free to search in any of the  $d$  directions, while all subsequent vertices are assigned types, and thus expand like a branching process. Thus, for each  $i = 1, \dots, d$ ,  $S_1^{(i)}$  is distributed as a Binomial( $a_i n, \frac{\lambda}{n}$ ) random variable. We can thus bound this step of the process:

$$\begin{aligned} \mathbb{P}(\langle \mathbf{S}_1 - \mathbf{S}_0, \boldsymbol{\mu} \rangle \geq (d+2) \log n) &= \mathbb{P}\left(e^{\langle \mathbf{S}_1 - \mathbf{S}_0, \boldsymbol{\mu} \rangle} \geq e^{(d+2) \log n}\right) \\ &\leq n^{-(d+2)} \mathbb{E} e^{\langle \mathbf{S}_1 - \mathbf{S}_0, \boldsymbol{\mu} \rangle} \\ &\leq n^{-(d+2)} \exp \left[ \sum_{i=1}^d a_i \lambda (e^{\mu_i} - 1) - \mu_{\min} \right] \\ &= O\left(n^{-(d+2)}\right) \end{aligned} \quad (17)$$

We only need this upper bound, since the first step of  $\mathbf{S}_t$  is stochastically bounded below by any subsequent step, so inequality (15) still holds. We now apply Lemma 3.2 to  $\mathbf{S}_t$  at time  $s = \beta \log n$

with  $x = 2$  and  $y = \frac{1}{2}$  (and use the inequality in (17) to handle the first step) to yield

$$\mathbb{P}(\langle \mathbf{S}_s - \mathbf{S}_0, \boldsymbol{\mu} \rangle \geq [2\beta(\rho - 1)\mu_{\max} + (d + 2)] \log n) \leq n^{-\eta_1\beta} + O(n^{-(d+2)})$$

$$\mathbb{P}\left(\langle \mathbf{S}_s - \mathbf{S}_0, \boldsymbol{\mu} \rangle \leq \frac{1}{2}\beta(\rho - 1)\mu_{\min} \log n\right) \leq n^{-\eta_2\beta}.$$

If we choose  $\beta > 0$  such that  $\eta_1\beta > d+1$ ,  $\eta_2\beta > d+1$  and  $\beta > \frac{2\mu_{\max}(d+2)}{\mu_{\min}(\rho-1)\gamma_2}$  (recall that  $m = \frac{d+1}{\gamma_2} \log n$ ), and we let  $K = \frac{2\beta(\rho-1)\mu_{\max}+(d+2)}{\mu_{\min}} + \beta + 1$ , then

$$\mathbb{P}(\|\mathbf{S}_s\|_1 + s \geq K \log n) \leq n^{-(d+1)} \quad (18)$$

$$\mathbb{P}\left(\|\mathbf{S}_s\|_1 \leq \frac{d+2}{\gamma_2} \log n\right) \leq n^{-(d+1)}. \quad (19)$$

To obtain the above inequalities, we used the estimates  $\langle \mathbf{S}_s - \mathbf{S}_0, \boldsymbol{\mu} \rangle \geq \mu_{\min} \|\mathbf{S}_s\|_1 - \|\mathbf{S}_0\|_{\infty}$ ,  $\langle \mathbf{S}_s - \mathbf{S}_0, \boldsymbol{\mu} \rangle \leq \mu_{\max} \|\mathbf{S}_s\|_1$  and  $\|\mathbf{S}_0\|_{\infty} / \mu_{\min} \leq \log n$  for  $n$  sufficiently large ( $\|\mathbf{S}_0\|_{\infty} = 1$ ). Equation (18) implies the second part of the lemma, since  $\|\mathbf{A}_s\|_1 \leq \|\mathbf{S}_s\|_1$ .

When  $\|\mathbf{S}_s\|_1 + s \leq K \log n$ , we have:

$$|U_s \cap \mathcal{N}^i(\mathbf{v})| \geq a_i n - K \log n \quad (20)$$

for  $\mathbf{v}$  of type  $j \neq i$ . This is because  $|A_s| \leq \|\mathbf{S}_s\|_1$ ,  $|R_s| \leq s$  (with equality here if  $\mathcal{A}_t$  is still alive at time  $s$ ), and the maximum number of neighbors that  $A_s \cup R_s$  can have in  $\mathcal{N}^i(\mathbf{v})$  is  $|A_s \cup R_s|$  as long as  $\mathbf{v}$  is not of type  $i$ . Therefore, provided  $\mathbf{A}_s \neq \mathbf{0}$ , the number of births that occur in  $\mathbf{S}_t$  which are lost in  $\mathbf{A}_t$  for  $t \leq s$  is at most

$$\sum_{t=0}^{s-1} \sum_{i=1}^d \sum_{k=1}^{a_i n - |U_t \cap \mathcal{N}^i(\mathbf{v}_t)|} \eta_k^{(t,i)} \leq Y^{(s)} \quad (21)$$

where  $Y^{(s)} \sim \text{Binomial}(d\beta K(\log n)^2, \frac{\lambda}{n})$ . So, conditioned on  $\|\mathbf{S}_s\|_1 + s \leq K \log n$ , we can bound the difference between  $\mathbf{S}_s$  and  $\mathbf{A}_s$  by a constant:

$$\begin{aligned} \mathbb{P}(\|\mathbf{A}_s\|_1 > 0, \|\mathbf{S}_s - \mathbf{A}_s\|_1 \geq d+2) &\leq \mathbb{P}(Y^{(s)} \geq d+2) \\ &= \sum_{k=d+2}^{d\beta K(\log n)^2} \binom{d\beta K(\log n)^2}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{d\beta K(\log n)^2 - k} \\ &\leq \sum_{k=d+2}^{d\beta K(\log n)^2} \left(\frac{\lambda d\beta K(\log n)^2}{n}\right)^k \\ &= O(n^{-(d+1)}). \end{aligned} \quad (22)$$

Finally, combining inequalities (18), (19), and (22) yields

$$\begin{aligned} \mathbb{P}\left(0 < \|\mathbf{A}_s\|_1 \leq \frac{d+2}{\gamma_2} \log n - (d+2)\right) &= O(n^{-(d+1)}), \\ \mathbb{P}\left(0 < \|\mathbf{A}_s\|_1 \leq \frac{d+1}{\gamma_2} \log n\right) &= O(n^{-(d+1)}), \\ \mathbb{P}(0 < \|\mathbf{A}_s\|_1 \leq m) &= O(n^{-(d+1)}). \end{aligned}$$

□

This means that, with high probability, at time  $s$  the process has either died out or there are at least  $m$  active vertices (in  $A_s$ ).

### 3.3 Merging Clusters of Size $r = n^{d-4/3}$

At this point, with high probability, the cluster-discovering algorithm started at a given vertex,  $\mathbf{v}$ , has either died out by time  $s = \beta \log n$ , or will continue to survive until at least time  $r = n^{d-4/3}$ . If the process has died out, then the size of the cluster containing  $\mathbf{v}$  is at most  $\beta \log n$ . What we now intend to show is that, if the cluster containing the vertex  $\mathbf{v}_1$  and the cluster containing the vertex  $\mathbf{v}_2$  each have at least  $r$  vertices, then  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are in the same connected component of the random site subgraph with high probability.

The clusters containing  $\mathbf{v}_1$  and  $\mathbf{v}_2$  will be generated as follows (we use  $\mathbf{v}_{1/2}$  to denote “ $\mathbf{v}_1$  or  $\mathbf{v}_2$ , respectively”, and likewise for similar notation):

1. Start the cluster-discovering algorithm,  $(R_t, A_t, U_t)$ , with parameter  $p$  at  $R_0 = \emptyset$ ,  $A_0 = \mathbf{v}_{1/2}$  and  $U_0 = V \setminus A_0$ . Continue until time  $s = \beta \log n$ .
2. Start the cluster-discovering algorithm  $(R'_t, A'_t, U'_t)$  with parameter  $p_1$  at  $R'_0 = R_s$ ,  $A'_0 = A_s$  and  $U'_0 = U_s$ . Continue until time  $r = n^{d-4/3}$ .
3. Let  $\Upsilon_{1/2} = R'_r \setminus R_s$ .

Note that, conditioned on the survival of each process,  $|\Upsilon_{1/2}| = r$ , and  $\Upsilon_{1/2} \subsetneq C_{\mathbf{v}_{1/2}}$ .

**Lemma 3.4.** *If  $\Upsilon_1$  and  $\Upsilon_2$  are defined for  $\mathbf{v}_1$  and  $\mathbf{v}_2$  as above and  $|\Upsilon_1| = |\Upsilon_2| = r$ , then:*

$$\mathbb{P}_{p_2}(C_{\mathbf{v}_1} \neq C_{\mathbf{v}_2}) = O(n^{-(d+1)})$$

*Proof of Lemma 3.4.* We will employ a sprinkling technique using the probability reserved in the first step, where the lower bounding process was defined using probability parameter  $p_1 = \lambda_1/n$ . Observe that we can write  $p = p_1 + p_2 - p_1 p_2$ , where  $p = \lambda/n$  and  $p_2$  is defined in this way. The random site subgraph of  $G$  with parameter  $p$ , call it  $G_p$ , can then be seen as  $G_p = \overline{G_{p_1} \cup G_{p_2}}$ , where  $G_{p_1}$  and  $G_{p_2}$  are random site subgraphs of  $G$  that are independently generated, the union is taken over their vertex sets, and the bar denotes the inclusion of all edges from  $G$  between vertices in  $G_{p_1} \cup G_{p_2}$ . We will further subdivide  $p_2$  in the same manner, but first it is crucial to note that:

$$p_2 = \frac{\lambda - \lambda_1}{n} + \frac{\lambda_1(\lambda - \lambda_1)}{n(n - \lambda_1)} = \frac{\epsilon}{n} + O(n^{-2}),$$

where  $\epsilon = \lambda - \lambda_1$ . We can now write

$$p_2 = \sum_{i=1}^{2d-4} p'_i - \sum_{1 \leq i < j \leq 2d-4} p'_i p'_j + \cdots + (-1)^{2d-3} \prod_{i=1}^{2d-4} p'_i \quad (23)$$

where  $p'_i = \epsilon_i/n + O(n^{-2})$  and  $\sum_i \epsilon_i = \epsilon$ . This is done merely by repeatedly subdividing  $G_p$  into a union of independent subgraphs as described above. Also, for the sprinkling argument to work as intended (so we can avoid dependencies), we will consider only the vertices in  $U_s$ , since the vertices in  $V \setminus U_s$  have already been fully considered for inclusion in  $G_p$ . Since  $|U_s \cap \mathcal{N}^i(\mathbf{v})| \geq a_i n - K \log n$  whenever  $\mathbf{v} \in U_s$  (see inequality (20)), this restriction will not significantly affect our estimates below.

We will first treat the cases  $d = 2, 3$  separately, then use an induction argument to prove the Lemma for  $d \geq 4$ .

**Case 1:**  $d = 2$ . We first observe that with high probability no row or column ( $\mathcal{P}_{(i)}^{(\ell)}$  for  $i \in \{1, 2\}$  and  $\ell \in \{1, \dots, a_i n\}$ ) contains more than  $\log^2 n$  vertices in  $\Upsilon_{1/2}$  because:

$$\begin{aligned} \mathbb{P} \left( \left| \mathcal{P}_{(i)}^{(\ell)} \cap \Upsilon_{1/2} \right| \geq \log^2 n \text{ for some } (i) \text{ and } (\ell) \right) &\leq \mathbb{P} \left( \left| \mathcal{P}_{(i)}^{(\ell)} \cap G_p \right| \geq \log^2 n \text{ for some } (i) \text{ and } (\ell) \right) \\ &\leq 2a_1 n \left[ e^{-\log^2 n} \sum_{k=1}^{a_1 n} \binom{a_1 n}{k} p^k e^k (1-p)^{a_1 n - k} \right] \\ &\leq 2a_1 n \exp \left[ -\log^2 n + (e-1)a_1 \lambda \right] \\ &= O(n^{1-\log n}). \end{aligned}$$

If no column ( $\mathcal{P}_{(1)}^{(\ell)}$  for  $\ell \in \{1, \dots, a_1 n\}$ ) has more than  $\log^2 n$  vertices that are in  $\Upsilon_{1/2}$ , then by the Pigeonhole Principle there are at least  $r/\log^2 n$  columns with at least one vertex in  $\Upsilon_{1/2}$ . Thus we have:

$$\mathbb{P} \left( \left| \mathcal{P}_{(1)}^{(\ell)} \cap \mathcal{N}^1(\Upsilon_{1/2}) \right| \leq \frac{n^{2/3}}{\log^2 n} \right) \leq O(n^{1-\log n}). \quad (24)$$

We denote by  $B_{p_2}(\Upsilon_{1/2})$  the set of vertices in  $\mathcal{N}^1(\Upsilon_{1/2})$  that are independently made occupied with parameter  $p_2$ . Note that  $B_{p_2}(\Upsilon_{1/2})$  contains independent copies of any vertices that have been previously observed. If there is an edge in  $G$  between  $\Upsilon_1$  and  $\Upsilon_2$  then  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are in the same component and we are finished. Otherwise, we seek to bound the probability that there is no edge in  $G$  between  $B_{p_2}(\Upsilon_1)$  and  $B_{p_2}(\Upsilon_2)$ . If there is a column that contains a vertex from each of these two sets then the two sets are connected. Let  $E_\ell$  be the event that column  $\ell$  ( $\mathcal{P}_{(1)}^{(\ell)}$ ) contains a vertex from  $B_{p_2}(\Upsilon_1)$  and  $B_{p_2}(\Upsilon_2)$ , and let  $E$  be the disjoint union of these events:

$$\begin{aligned} E_\ell &= \{ \mathcal{P}_{(1)}^{(\ell)} \cap B_{p_2}(\Upsilon_1) \neq \emptyset \} \cap \{ \mathcal{P}_{(1)}^{(\ell)} \cap B_{p_2}(\Upsilon_2) \neq \emptyset \}, \\ E &= \bigsqcup_{\ell=1}^{a_1 n} E_\ell, \end{aligned}$$

We wish to bound the likelihood of the complementary event:

$$\begin{aligned} \mathbb{P}(E_\ell^c) &\leq 1 - \left[ 1 - (1-p_2)^{n^{2/3}/\log^2 n} \right]^2 \\ &\leq 1 - \left[ 1 - \exp \left( -\frac{\epsilon}{n^{1/3} \log^2 n} \right) \right]^2 \\ &\leq 1 - \left[ \frac{\epsilon}{n^{1/3} \log^2 n} - \frac{\epsilon^2}{2n^{2/3} \log^4 n} \right]^2 \\ &\lesssim 1 - \frac{\epsilon^2}{2n^{2/3} \log^4 n} \\ &\leq \exp \left( -\frac{\epsilon^2}{2n^{2/3} \log^4 n} \right). \end{aligned}$$

In the second and fifth lines above we have applied the bound  $1-x \leq e^{-x}$ , in the third line we used  $1-e^{-x} \geq x - x^2/2$  for  $x \geq 0$ , and the fourth line bounds the third for sufficiently large  $n$  (which we

denote by  $\lesssim$ ). Because the events  $E_\ell$  depend on disjoint columns of vertices, they are independent, so we have:

$$\begin{aligned}\mathbb{P}(E^c) &= \mathbb{P}\left(\bigcap_{\ell=1}^{a_1 n} E_\ell^c\right) \\ &= \prod_{\ell=1}^{a_1 n} \mathbb{P}(E_\ell^c) \\ &\lesssim \exp\left(-\frac{a_1 \epsilon^2 n^{1/3}}{2 \log^4 n}\right).\end{aligned}\tag{25}$$

So, combining (24) and (25), for  $d = 2$  the probability that  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are not in the same component is bounded above by  $O(n^{1-\log n})$ .

**Case 2:**  $d = 3$ . We would like to begin by sprinkling vertices in orthogonal directions from each of the sets  $\Upsilon_1$  and  $\Upsilon_2$ , but we would like to sprinkle in the directions that will give the greatest likelihood of adjoining the two sets. For this, we need the following fact about subsets of  $\mathbb{Z}^3$ .

**Lemma 3.5.** *Let  $A \subset \mathbb{Z}^3$ , and let  $A_x$ ,  $A_y$ , and  $A_z$ , respectively, be projections of  $A$  onto the coordinate planes,  $yz$ -plane,  $xz$ -plane, and  $xy$ -plane. Then*

$$|A|^{2/3} \leq |A_x|^{1/3} |A_y|^{1/3} |A_z|^{1/3}.$$

Lemma 3.5 is a restatement of the main theorem of [11] for  $d = 3$ , and it implies that at least one of the projections has size at least  $|A|^{2/3}$ . Applying this to the set  $\Upsilon_1$ , suppose without loss of generality (since the constants  $a_i$  will not play a significant part here) that  $|\mathcal{P}_{(3)}^{(\ell_3)} \cap \mathcal{N}^3(\Upsilon_1)| \geq r^{2/3}$  for all  $\ell_3$ . Throughout this section, we will also condition on the complement of the event in Lemma 3.1, which implies that no two-dimensional plane orthogonal to one of the coordinate axes contains more than  $n^{2/3} \log n$  vertices in  $\Upsilon_2$ . In particular, this implies that  $|\mathcal{P}_{(2,3)}^{(\ell_2, \ell_3)} \cap \Upsilon_2| \leq C n^{2/3} \log n$  for all  $(\ell_2, \ell_3)$ . This bounds the number of points that can map to the same point when the set  $\Upsilon_2$  is projected onto the span of  $\mathbf{e}_2$  and  $\mathbf{e}_3$ , so the Pigeonhole Principle implies that  $|\mathcal{P}_{(1)}^{(\ell_1)} \cap \mathcal{N}^1(\Upsilon_2)| \geq \frac{n}{C \log n}$ .

We begin by sprinkling occupied vertices independently in the neighborhood  $\mathcal{N}^3(\Upsilon_1)$  with probability  $p'_1 = \epsilon_1/n$  and in the neighborhood  $\mathcal{N}^1(\Upsilon_2)$  with probability  $p'_2 = \epsilon_2/n$ . We will refer to the sets of newly added vertices (including those which may have already been present in  $G_{p_1}$ ) as  $B_{p'_1}^3(\Upsilon_1)$  and  $B_{p'_2}^1(\Upsilon_2)$ , respectively. The key observation here is that  $\mathbf{v}_1$  and  $\mathbf{v}_2$  will be in the same component if  $\mathcal{N}^2(B_{p'_1}^3(\Upsilon_1)) \cap \mathcal{N}^2(B_{p'_2}^1(\Upsilon_2)) \neq \emptyset$ . That is, if there exists a pair  $(\ell_1, \ell_3)$  such that  $\mathcal{P}_{(1,3)}^{(\ell_1, \ell_3)}$  contains a vertex in  $B_{p'_1}^3(\Upsilon_1)$  and a vertex in  $B_{p'_2}^1(\Upsilon_2)$ , then  $\mathbf{v}_1$  and  $\mathbf{v}_2$  must be in the same component. Denote by  $E_{(\ell_1, \ell_3)}$  the event that  $\mathcal{P}_{(1,3)}^{(\ell_1, \ell_3)}$  contains a vertex in  $B_{p'_1}^3(\Upsilon_1)$  and a vertex in  $B_{p'_2}^1(\Upsilon_2)$ , and let  $E$  be the disjoint union of these events over all pairs  $(\ell_1, \ell_3)$ :

$$\begin{aligned}E_{(\ell_1, \ell_3)} &= \{\mathcal{P}_{(1,3)}^{(\ell_1, \ell_3)} \cap B_{p'_1}^3(\Upsilon_1) \neq \emptyset\} \cap \{\mathcal{P}_{(1,3)}^{(\ell_1, \ell_3)} \cap B_{p'_2}^1(\Upsilon_2) \neq \emptyset\}, \\ E &= \bigsqcup_{\ell_1=1}^{a_1 n} \bigsqcup_{\ell_3=1}^{a_3 n} E_{(\ell_1, \ell_3)}.\end{aligned}$$

We now define the following notation:

$$k_{\ell_1}^{(1)} := \# \left\{ v_2 \in \{1, \dots, a_2 n\} : \mathbf{v} \in \mathcal{P}_{(1)}^{(\ell_1)} \cap \Upsilon_1 \right\},$$

$$k_{\ell_3}^{(2)} := \# \left\{ v_2 \in \{1, \dots, a_2 n\} : \mathbf{v} \in \mathcal{P}_{(3)}^{(\ell_3)} \cap \Upsilon_2 \right\}.$$

In other words,  $k_{\ell_1}^{(1)}$  is the number of columns,  $\mathcal{P}_{(1,2)}^{(\ell_1, \ell_2)} \subset \mathcal{P}_{(1)}^{(\ell_1)}$ , which contain at least one vertex in  $\Upsilon_1$ . Similarly,  $k_{\ell_3}^{(2)}$  is the number of rows,  $\mathcal{P}_{(2,3)}^{(\ell_2, \ell_3)} \subset \mathcal{P}_{(3)}^{(\ell_3)}$ , which contain at least one vertex in  $\Upsilon_2$  (see Figure 1). Note that  $k_{\ell_1}^{(1)} = O(n^{2/3} \log n)$  and  $k_{\ell_3}^{(2)} = O(n^{2/3} \log n)$ , since we are conditioning on the complement of the event in Lemma 3.1.

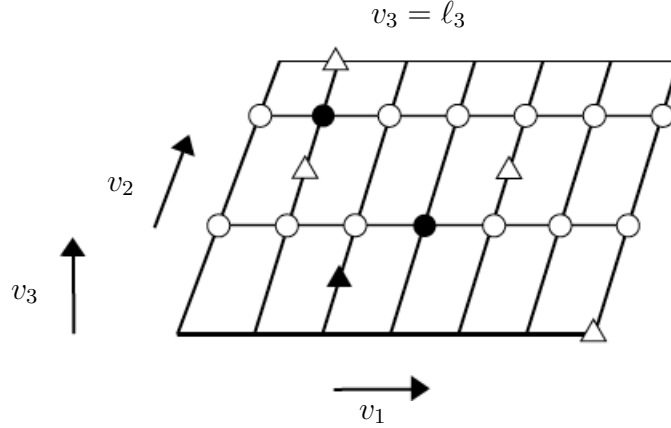


Figure 1: Solid triangles and circles are vertices in  $\Upsilon_1$  and  $\Upsilon_2$ , respectively. Empty triangles are neighbors of vertices in  $\Upsilon_1$  in the  $v_3$  direction, and empty circles are neighbors of vertices in  $\Upsilon_2$  in the  $v_1$  direction. Then  $k_{\ell_1}^{(1)}$  is the number of triangles (solid or empty) in the column  $v_1 = \ell_1$  (e.g.  $k_0^{(1)} = 0$ ,  $k_1^{(1)} = 2$ ), and  $k_{\ell_3}^{(2)}$  is the number of rows with circles in the plane  $v_3 = \ell_3$  (here  $k_{\ell_3}^{(2)} = 2$ ). Also,  $E_{(\ell_1, \ell_3)}$  is the event that in the column  $v_1 = \ell_1$ , at least one triangle (solid or empty) is made occupied with parameter  $p_1'$  and at least one circle (solid or empty) is made occupied with parameter  $p_2'$ .

We can now write the probabilities of  $E_{(\ell_1, \ell_3)}^c$  and  $E^c$ , conditional on the values of  $k_{\ell_1}^{(1)}$  and  $k_{\ell_3}^{(2)}$ , as:

$$\mathbb{P} \left( E_{(\ell_1, \ell_3)}^c \right) = 1 - \left( 1 - \left( 1 - \frac{\epsilon_1}{n} \right)^{k_{\ell_1}^{(1)}} \right) \left( 1 - \left( 1 - \frac{\epsilon_2}{n} \right)^{k_{\ell_3}^{(2)}} \right),$$

$$\mathbb{P} (E^c) = \prod_{\ell_1=1}^{a_1 n} \prod_{\ell_3=1}^{a_3 n} \left[ 1 - \left( 1 - \left( 1 - \frac{\epsilon_1}{n} \right)^{k_{\ell_1}^{(1)}} \right) \left( 1 - \left( 1 - \frac{\epsilon_2}{n} \right)^{k_{\ell_3}^{(2)}} \right) \right],$$

and we wish to bound these quantities. To this end, we observe that for large  $n$ :

$$\begin{aligned}
1 - \left(1 - \frac{\epsilon_{1/2}}{n}\right)^k &= \frac{k\epsilon_{1/2}}{n} - \sum_{i=2}^k \binom{k}{i} \left(\frac{\epsilon_{1/2}}{n}\right)^i (-1)^i \\
&\geq \frac{k\epsilon_{1/2}}{n} - \left(\frac{k\epsilon_{1/2}}{n}\right)^2 \sum_{i=0}^k \left(\frac{k\epsilon_{1/2}}{n}\right)^i \\
&\gtrsim \frac{k\epsilon_{1/2}}{2n},
\end{aligned} \tag{26}$$

provided  $k = o(n)$ , which is the case for  $k_{\ell_1}^{(1)}$  and  $k_{\ell_3}^{(2)}$ . Applying this inequality and  $1 - x \leq e^{-x}$  gives:

$$\begin{aligned}
\mathbb{P}(E^c) &\lesssim \prod_{\ell_1=1}^{a_1 n} \prod_{\ell_3=1}^{a_3 n} \left[ 1 - \left(\frac{k_{\ell_1}^{(1)} \epsilon_1}{2n}\right) \left(\frac{k_{\ell_3}^{(2)} \epsilon_2}{2n}\right) \right] \\
&\leq \exp \left[ -\frac{\epsilon_1 \epsilon_2}{4n^2} \sum_{\ell_1=1}^{a_1 n} k_{\ell_1}^{(1)} \sum_{\ell_3=1}^{a_3 n} k_{\ell_3}^{(2)} \right].
\end{aligned} \tag{27}$$

What remains now is to obtain lower bounds for the sums in the above expression. It follows from the definitions of  $k_{\ell_1}^{(1)}$  and  $k_{\ell_3}^{(2)}$  that for any  $\ell_3$

$$\sum_{\ell_1=1}^{a_1 n} k_{\ell_1}^{(1)} = \left| \mathcal{P}_{(3)}^{(\ell_3)} \cap \mathcal{N}^3(\Upsilon_1) \right| \geq r^{2/3} = n^{10/9},$$

and for any  $\ell_1$

$$\sum_{\ell_3=1}^{a_3 n} k_{\ell_3}^{(2)} = \left| \mathcal{P}_{(1)}^{(\ell_1)} \cap \mathcal{N}^1(\Upsilon_2) \right| \geq \frac{n}{C \log n}.$$

Applying the last two lower bounds to (27), then handling the exceptional case where some plane contains more than  $Cn^{2/3} \log n$  vertices in  $\Upsilon_{1/2}$  by using Lemma 3.1, yields

$$\begin{aligned}
\mathbb{P}(E^c) &\lesssim \exp \left[ -\frac{\epsilon_1 \epsilon_2}{4n^2} \left( \frac{n}{C \log n} \right) (n^{10/9}) \right] + O(n^{-(d+1)}) \\
&= O(n^{-(d+1)}).
\end{aligned}$$

**Case 3:**  $d \geq 4$ . We will ultimately use the same argument here as in Case 2, but first we will sprinkle vertices in the directions  $\mathbf{e}_d, \mathbf{e}_{d-1}, \dots, \mathbf{e}_4$  sequentially. At that point we will have  $\prod_{i=4}^d a_i n^{d-3}$  independent 3-dimensional Hamming Tori in which the two sets can connect. So that we can independently connect vertices to the sets  $\Upsilon_{1/2}$ , we will write  $p_2$  as in equation (23) with  $k = 2d - 4$ . We will denote by  $B_p^i(J)$  the set of vertices that are independently made occupied with probability  $p$  in the neighborhood  $\mathcal{N}^i(J)$ .

We now begin by sprinkling vertices in the  $\mathbf{e}_d$  direction to generate the sets  $B_{p_{1/2}}^d(\Upsilon_{1/2})$ . By conditioning on the complement of the event in Lemma 3.1, we have that for every  $(\ell_1, \dots, \ell_{d-1})$

$$\left| \mathcal{P}_{(1, \dots, d-1)}^{(\ell_1, \dots, \ell_{d-1})} \cap \Upsilon_{1/2} \right| \leq Cn^{2/3} (\log n)^{d-2}.$$

This bound says that when the sets  $\Upsilon_{1/2}$  are projected onto the span of  $\mathbf{e}_1, \dots, \mathbf{e}_{d-1}$ , no more than this many vertices can be mapped onto the same point. Thus, by the Pigeonhole Principle, we have a lower bound on the size of the projection:

$$\left| \mathcal{P}_{(d)}^{(\ell_d)} \cap \mathcal{N}^d(\Upsilon_{1/2}) \right| \geq \frac{r}{C n^{2/3} (\log n)^{d-2}} = \frac{n^{d-2}}{C (\log n)^{d-2}}$$

for any  $\ell_d$ . Conditional on this bound, we obtain the following bound on the number of newly occupied vertices in  $B_{p'_{1/2}}^d(\Upsilon_{1/2})$  in a fixed  $(d-1)$ -dimensional Hamming torus:

$$\begin{aligned} \mathbb{P} \left( \left| \mathcal{P}_{(d)}^{(\ell_d)} \cap B_{p'_{1/2}}^d(\Upsilon_{1/2}) \right| \leq \frac{n^{d-3}}{(\log n)^{d-1}} \right) &\leq \exp \left[ \frac{n^{d-3}}{(\log n)^{d-1}} \right] \cdot \mathbb{E} \exp \left[ - \left| \mathcal{P}_{(d)}^{(\ell_d)} \cap B_{p'_{1/2}}^d(\Upsilon_{1/2}) \right| \right] \\ &= \exp \left[ \frac{n^{d-3}}{(\log n)^{d-1}} \right] \cdot \left[ 1 - \frac{\epsilon_{1/2}}{n} (1 - e^{-1}) \right]^{\left| \mathcal{P}_{(d)}^{(\ell_d)} \cap \mathcal{N}^d(\Upsilon_{1/2}) \right|} \\ &\leq \exp \left[ \frac{n^{d-3}}{(\log n)^{d-1}} - \frac{\epsilon_{1/2} n^{d-3}}{C (\log n)^{d-2}} (1 - e^{-1}) \right] \\ &\lesssim \exp \left[ -\eta'_3 \frac{n^{d-3}}{(\log n)^{d-2}} \right] \end{aligned}$$

for some constant  $\eta'_3 > 0$ . In the first line above we exponentiated and applied Markov's inequality. In the second line we used the moment generating function for a binomial random variable, and in the third line we used the bound on the size of the projection that we obtained above and the inequality  $1 - x \leq e^{-x}$ . By the union bound, we have the following bound on the number of newly occupied vertices in each  $(d-1)$ -dimensional Hamming Torus:

$$\begin{aligned} \mathbb{P} \left( \exists \ell_d \text{ s.t. } \left| \mathcal{P}_{(d)}^{(\ell_d)} \cap B_{p'_{1/2}}^d(\Upsilon_{1/2}) \right| \leq \frac{n^{d-3}}{(\log n)^{d-1}} \right) &\leq a_d n \cdot \exp \left[ -\eta'_3 \frac{n^{d-3}}{(\log n)^{d-2}} \right] \\ &\lesssim \exp \left[ -\eta_3 \frac{n^{d-3}}{(\log n)^{d-2}} \right] \end{aligned} \quad (28)$$

for some constant  $\eta_3 > 0$ . For  $d = 4$  this is as far as we will need to project before applying the argument from Case 2. For that argument to work, and so we can continue to project down to lower dimensions for  $d > 4$ , we will need an estimate for the maximum number of vertices in  $\mathcal{P}_{(d)}^{(\ell_d)} \cap B_{p'_{1/2}}^d(\Upsilon_{1/2})$  that can map to a single vertex when the set is projected onto a plane of one less dimension. This will be easy by again using Markov's inequality and the moment generating function for a binomial random variable. For any  $1 \leq i_1 < i_2 < \dots < i_{d-2} < d$ :

$$\begin{aligned} &\mathbb{P} \left( \exists (\ell_{i_1}, \dots, \ell_{i_{d-2}}, \ell_d) \text{ s.t. } \left| \mathcal{P}_{(i_1, \dots, i_{d-2}, d)}^{(\ell_{i_1}, \dots, \ell_{i_{d-2}}, \ell_d)} \cap B_{p'_{1/2}}^d(\Upsilon_{1/2}) \right| \geq \log^2 n \right) \\ &\leq \prod_i a_i n^{d-1} \exp [-\log^2 n] \cdot \mathbb{E} \exp \left[ \left| \mathcal{P}_{(i_1, \dots, i_{d-2}, d)}^{(\ell_{i_1}, \dots, \ell_{i_{d-2}}, \ell_d)} \cap B_{p'_{1/2}}^d(\Upsilon_{1/2}) \right| \right] \\ &= \prod_i a_i n^{d-1} \exp [-\log^2 n] \cdot \left[ 1 + \frac{\epsilon_{1/2}}{n} (e - 1) \right]^{\left| \mathcal{P}_{(i_1, \dots, i_{d-2}, d)}^{(\ell_{i_1}, \dots, \ell_{i_{d-2}}, \ell_d)} \cap \mathcal{N}^d(\Upsilon_{1/2}) \right|} \\ &\leq \prod_i a_i n^{d-1} \exp \left[ -\log^2 n + \frac{\epsilon_{1/2} C n^{2/3} (\log n)^{d-2}}{n} (e - 1) \right] \\ &\lesssim \exp [-\eta_4 \log^2 n]. \end{aligned} \quad (29)$$

The third and fourth lines above are conditional on the complement of the event in Lemma 3.1, and assume that the sets  $\Upsilon_{1/2}$  have been determined accordingly.

Now, if  $d > 4$ , we will repeat the above argument until we are within 3-dimensional Hamming tori. Suppose, for some  $k$  ( $1 \leq k \leq d-4$ ), that

$$\left| \mathcal{P}_{(d-k+1, \dots, d)}^{(\ell_{d-k+1}, \dots, \ell_d)} \cap B_{P'_{(2k-1)/(2k)}}^{d-k+1} (B_{P'_{(2k-3)/(2k-2)}}^{d-k+2} (\dots (B_{P'_{1/2}}^d(\Upsilon_{1/2})) \dots) \right| \geq N \quad (30)$$

for all  $(\ell_{d-k+1}, \dots, \ell_d)$  where  $N = \omega(n \log^3 n)$ , and that

$$\left| \mathcal{P}_{(i_1, \dots, i_{d-k-1}, d-k+1, \dots, d)}^{(\ell_{i_1}, \dots, \ell_{i_{d-k-1}}, \ell_{d-k+1}, \dots, \ell_d)} \cap B_{P'_{(2k-1)/(2k)}}^{d-k+1} (B_{P'_{(2k-3)/(2k-2)}}^{d-k+2} (\dots (B_{P'_{1/2}}^d(\Upsilon_{1/2})) \dots) \right| \leq \log^2 n \quad (31)$$

for all  $1 \leq i_1 < \dots < i_{d-k-1} < d-k+1$  and  $(\ell_{i_1}, \dots, \ell_{i_{d-k-1}}, \ell_{d-k+1}, \dots, \ell_d)$ . The first inequality above says that the sets  $\Upsilon_{1/2}$  each have at least  $N$  (most recently occupied) neighbors in each of the  $\Theta(n^k)$   $(d-k)$ -dimensional sub-Hamming tori after newly occupied vertices have been sprinkled successively in the directions  $\mathbf{e}_d, \dots, \mathbf{e}_{d-k+1}$ . The second inequality says that no line contains more than  $\log^2 n$  of the most recently added occupied vertices. Under these two assumptions, and by the Pigeonhole Principle, we have that the size of the projection onto the span of  $\mathbf{e}_1, \dots, \mathbf{e}_{d-k-1}$  is bounded by:

$$\left| \mathcal{P}_{(d-k, \dots, d)}^{(\ell_{d-k}, \dots, \ell_d)} \cap \mathcal{N}^{d-k} (B_{P'_{(2k-1)/(2k)}}^{d-k+1} (\dots (B_{P'_{1/2}}^d(\Upsilon_{1/2})) \dots) \right| \geq \frac{N}{\log^2 n}$$

for any  $(\ell_{d-k}, \dots, \ell_d)$ . Using Markov's inequality and the moment generating function for a binomial random variable as we did above, we have that

$$\begin{aligned} \mathbb{P} \left( \exists (\ell_{d-k}, \dots, \ell_d) \text{ s.t. } \left| \mathcal{P}_{(d-k, \dots, d)}^{(\ell_{d-k}, \dots, \ell_d)} \cap B_{P'_{(2k+1)/(2k+2)}}^{d-k} (B_{P'_{(2k-1)/(2k)}}^{d-k+1} (\dots (B_{P'_{1/2}}^d(\Upsilon_{1/2})) \dots) \right| \leq \frac{N}{n \log^3 n} \right) \\ \lesssim \exp \left[ -\eta_{2k+1} \frac{N}{n \log^2 n} \right] \end{aligned} \quad (32)$$

for some constant  $\eta_{2k+1} > 0$ . Similarly, for all  $1 \leq i_1 < \dots < i_{d-k-1} < d-k$  and all  $\ell = (\ell_{i_1}, \dots, \ell_{i_{d-k-2}}, \ell_{d-k}, \dots, \ell_d)$ ,

$$\left| \mathcal{P}_{(i_1, \dots, i_{d-k-2}, d-k, \dots, d)}^{(\ell_{i_1}, \dots, \ell_{i_{d-k-2}}, \ell_{d-k}, \dots, \ell_d)} \cap \mathcal{N}^{d-k} (B_{P'_{(2k-1)/(2k)}}^{d-k+1} (\dots (B_{P'_{1/2}}^d(\Upsilon_{1/2})) \dots) \right| \leq \log^2 n,$$

which implies that

$$\begin{aligned} \mathbb{P} \left( \exists \ell \text{ s.t. } \left| \mathcal{P}_{(i_1, \dots, i_{d-k-2}, d-k, \dots, d)}^{(\ell_{i_1}, \dots, \ell_{i_{d-k-2}}, \ell_{d-k}, \dots, \ell_d)} \cap B_{P'_{(2k+1)/(2k+2)}}^{d-k} (B_{P'_{(2k-1)/(2k)}}^{d-k+1} (\dots (B_{P'_{1/2}}^d(\Upsilon_{1/2})) \dots) \right| \geq \log^2 n \right) \\ \lesssim \exp \left[ -\eta_{2k+2} \log^2 n \right]. \end{aligned} \quad (33)$$

Putting together inequalities (28) through (33) inductively gives (for  $d \geq 4$ ):

$$\begin{aligned} \mathbb{P} \left( \exists \ell \text{ s.t. } \left| \mathcal{P}_{(4, \dots, d)}^{(\ell_4, \dots, \ell_d)} \cap B_{P'_{(2d-7)/(2d-6)}}^4 (B_{P'_{(2d-9)/(2d-8)}}^5 (\dots (B_{P'_{1/2}}^d(\Upsilon_{1/2})) \dots) \right| \leq \frac{n}{(\log n)^{4d-13}} \right) \\ \lesssim \exp \left[ -\eta_{2d-5} \log^2 n \right]. \end{aligned} \quad (34)$$

Here,  $\ell = (\ell_4, \dots, \ell_d)$ , and henceforth  $\ell$  will be the vector in the superscript of the next “ $\mathcal{P}$ ”. We used the union bound to control the probability above, and the constant  $\eta_{2d-5} > 0$  is chosen such

that the sum of the probabilities of the events involved is bounded asymptotically by the above expression. These events are still conditional on complement of the event in Lemma 3.1. Likewise, we have:

$$\begin{aligned} \mathbb{P} \left( \exists \ell \text{ s.t. } \left| \mathcal{P}_{(1,2,4,\dots,d)}^{(\ell_1,\ell_2,\ell_4,\dots,\ell_d)} \cap B_{p'_{(2d-7)}}^4 (B_{p'_{(2d-9)}}^5 (\dots (B_{p'_1}^d(\Upsilon_1)) \dots) \right| \geq \log^2 n \right) \\ \lesssim \exp \left[ -\eta_{2d-4} \log^2 n \right], \end{aligned} \quad (35)$$

and

$$\begin{aligned} \mathbb{P} \left( \exists \ell \text{ s.t. } \left| \mathcal{P}_{(2,3,4,\dots,d)}^{(\ell_2,\ell_3,\ell_4,\dots,\ell_d)} \cap B_{p'_{(2d-6)}}^4 (B_{p'_{(2d-8)}}^5 (\dots (B_{p'_2}^d(\Upsilon_2)) \dots) \right| \geq \log^2 n \right) \\ \lesssim \exp \left[ -\eta_{2d-3} \log^2 n \right]. \end{aligned} \quad (36)$$

Within each 3-dimensional Hamming torus,  $\mathcal{P}_{(4,\dots,d)}^{(\ell_4,\dots,\ell_d)}$ , we can now apply the argument from Case 2. This time we will not need Lemma 3.5 because we have a better bound on the number of (most recently) occupied vertices in any line. Fixing  $(\ell_4, \dots, \ell_d)$ , if we project the vertices most recently added to  $\Upsilon_1$  onto the plane spanned by  $\mathbf{e}_1$  and  $\mathbf{e}_2$ , then inequalities (34) and (35) imply (by the Pigeonhole Principle) that with high probability

$$\left| \mathcal{P}_{(3,\dots,d)}^{(\ell_3,\dots,\ell_d)} \cap \mathcal{N}^3(B_{p'_{(2d-7)}}^4 (\dots (B_{p'_1}^d(\Upsilon_1)) \dots) \right| \geq \frac{n}{(\log n)^{4d-11}}. \quad (37)$$

Likewise, if we project the vertices added most recently to  $\Upsilon_2$  onto the plane spanned by  $\mathbf{e}_2$  and  $\mathbf{e}_3$  and apply the Pigeonhole Principle with inequalities (34) and (36), then the size of the projection within each 3-dimensional sub-Hamming torus is at least:

$$\left| \mathcal{P}_{(1,4,\dots,d)}^{(\ell_1,\ell_4,\dots,\ell_d)} \cap \mathcal{N}^1(B_{p'_{(2d-6)}}^4 (\dots (B_{p'_2}^d(\Upsilon_2)) \dots) \right| \geq \frac{n}{(\log n)^{4d-11}}. \quad (38)$$

Now, as in Case 2, we will condition on the last two inequalities, and sprinkle new vertices independently in the  $\mathbf{e}_3$  direction from  $B_{p'_{(2d-7)}}^4 (\dots (B_{p'_1}^d(\Upsilon_1)) \dots)$  and in the  $\mathbf{e}_1$  direction from  $B_{p'_{(2d-6)}}^4 (\dots (B_{p'_2}^d(\Upsilon_2)) \dots)$ . The key observation here is that  $\mathbf{v}_1$  and  $\mathbf{v}_2$  will be in the same component if

$$\mathcal{N}^2(B_{p'_{(2d-5)}}^3 (\dots (B_{p'_1}^d(\Upsilon_1)) \dots) \cap \mathcal{N}^2(B_{p'_{(2d-4)}}^1 (\dots (B_{p'_2}^d(\Upsilon_2)) \dots) \neq \emptyset.$$

That is, if there exists a vector  $(\ell_1, \ell_3, \ell_4, \dots, \ell_d)$  such that  $\mathcal{P}_{(1,3,4,\dots,d)}^{(\ell_1,\ell_3,\ell_4,\dots,\ell_d)}$  contains a vertex in  $B_{p'_{(2d-5)}}^3 (\dots (B_{p'_1}^d(\Upsilon_1)) \dots)$  and a vertex in  $B_{p'_{(2d-4)}}^1 (\dots (B_{p'_2}^d(\Upsilon_2)) \dots)$ . If we define the following events

$$E_{(\ell_1,\ell_3,\ell_4,\dots,\ell_d)} := \left\{ B_{p'_{(2d-5)}}^3 (\dots (B_{p'_1}^d(\Upsilon_1)) \dots) \cap B_{p'_{(2d-4)}}^1 (\dots (B_{p'_2}^d(\Upsilon_2)) \dots) \cap \mathcal{P}_{(1,3,4,\dots,d)}^{(\ell_1,\ell_3,\ell_4,\dots,\ell_d)} \neq \emptyset \right\}$$

$$E_{(\ell_4,\dots,\ell_d)} := \bigcup_{\ell_1=1}^{a_1 n} \bigcup_{\ell_3=1}^{a_3 n} E_{(\ell_1,\ell_3,\ell_4,\dots,\ell_d)}$$

$$E := \bigcup_{\ell_4=1}^{a_4 n} \dots \bigcup_{\ell_d=1}^{a_d n} E_{(\ell_4,\dots,\ell_d)}$$

then  $E$  is contained in the event that  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are in the same component. To estimate the probabilities of these events, let us temporarily fix  $(\ell_4, \dots, \ell_d)$ , and define the variables:

$$\begin{aligned} k_{\ell_1}^{(1)} &:= \# \left\{ v_2 \in \{1, \dots, a_2 n\} : \mathbf{v} \in \mathcal{P}_{(1,4,\dots,d)}^{(\ell_1, \ell_4, \dots, \ell_d)} \cap B_{p'_{(2d-7)}}^4 (\dots (B_{p'_1}^d(\Upsilon_1)) \dots) \right\}, \\ k_{\ell_3}^{(2)} &:= \# \left\{ v_2 \in \{1, \dots, a_2 n\} : \mathbf{v} \in \mathcal{P}_{(3,4,\dots,d)}^{(\ell_3, \ell_4, \dots, \ell_d)} \cap B_{p'_{(2d-6)}}^4 (\dots (B_{p'_2}^d(\Upsilon_2)) \dots) \right\}. \end{aligned}$$

In other words,  $k_{\ell_1}^{(1)}$  is the number of columns,  $\mathcal{P}_{(1,2,4,\dots,d)}^{(\ell_1, \ell_2, \ell_4, \dots, \ell_d)} \subset \mathcal{P}_{(1,4,\dots,d)}^{(\ell_1, \ell_4, \dots, \ell_d)}$ , which contain at least one vertex in  $B_{p'_{(2d-7)}}^4 (\dots (B_{p'_1}^d(\Upsilon_1)) \dots)$ . Similarly,  $k_{\ell_3}^{(2)}$  is the number of rows,  $\mathcal{P}_{(2,3,4,\dots,d)}^{(\ell_2, \ell_3, \ell_4, \dots, \ell_d)} \subset \mathcal{P}_{(3,4,\dots,d)}^{(\ell_3, \ell_4, \dots, \ell_d)}$ , which contain at least one vertex in  $B_{p'_{(2d-6)}}^4 (\dots (B_{p'_2}^d(\Upsilon_2)) \dots)$ . We can now estimate the probability of  $E^c$ , conditional on the complements of the events in inequalities (34) - (36).

$$\begin{aligned} \mathbb{P} \left( E_{(\ell_1, \ell_3, \ell_4, \dots, \ell_d)}^c \right) &= 1 - \left( 1 - \left( 1 - p'_{(2d-5)} \right)^{k_{\ell_1}^{(1)}} \right) \left( 1 - \left( 1 - p'_{(2d-4)} \right)^{k_{\ell_3}^{(2)}} \right) \\ \mathbb{P} \left( E_{(\ell_4, \dots, \ell_d)}^c \right) &\lesssim \prod_{\ell_1=1}^{a_1 n} \prod_{\ell_3=1}^{a_3 n} \left[ 1 - \left( 1 - \left( 1 - \frac{\epsilon(2d-5)}{2n} \right)^{k_{\ell_1}^{(1)}} \right) \left( 1 - \left( 1 - \frac{\epsilon(2d-4)}{2n} \right)^{k_{\ell_3}^{(2)}} \right) \right] \\ &\lesssim \prod_{\ell_1=1}^{a_1 n} \prod_{\ell_3=1}^{a_3 n} \left[ 1 - \frac{\epsilon(2d-5)\epsilon(2d-4)}{16n^2} k_{\ell_1}^{(1)} k_{\ell_3}^{(2)} \right] \\ &\leq \exp \left[ - \frac{\epsilon(2d-5)\epsilon(2d-4)}{16n^2} \sum_{\ell_1=1}^{a_1 n} k_{\ell_1}^{(1)} \sum_{\ell_3=1}^{a_3 n} k_{\ell_3}^{(2)} \right] \\ &\leq \exp \left[ - \frac{\epsilon(2d-5)\epsilon(2d-4)}{16n^2} \left( \frac{n}{(\log n)^{4d-11}} \right)^2 \right] \\ &= \exp \left[ - \frac{\epsilon(2d-5)\epsilon(2d-4)}{16} \cdot \frac{1}{(\log n)^{8d-22}} \right] \\ \mathbb{P} (E^c) &\lesssim \exp \left[ - \frac{\epsilon(2d-5)\epsilon(2d-4)}{16} \cdot \frac{\prod_{i=4}^d a_i n^{d-3}}{(\log n)^{8d-22}} \right] \tag{39} \end{aligned}$$

In the second line of the above inequalities, the probability on the left is asymptotically less than the quantity on the right because  $p'_i = \epsilon_i/n + O(n^2)$ . The third line is obtained from inequality (26), which holds here because  $k_{\ell_1}^{(1)}, k_{\ell_3}^{(2)} = o(n)$  with high probability from inequalities (35) and (36). The fourth line is an application of the bound  $1 - x \leq e^{-x}$ . The fifth line is from the definitions of  $k_{\ell_1}^{(1)}$  and  $k_{\ell_3}^{(2)}$ , the observation that

$$\begin{aligned} \sum_{\ell_1=1}^{a_1 n} k_{\ell_1}^{(1)} &= \left| \mathcal{P}_{(3,4,\dots,d)}^{(\ell_3, \ell_4, \dots, \ell_d)} \cap \mathcal{N}^3(B_{p'_{(2d-7)}}^4 (\dots (B_{p'_1}^d(\Upsilon_1)) \dots) \right|, \\ \sum_{\ell_3=1}^{a_3 n} k_{\ell_3}^{(2)} &= \left| \mathcal{P}_{(1,4,\dots,d)}^{(\ell_1, \ell_4, \dots, \ell_d)} \cap \mathcal{N}^1(B_{p'_{(2d-6)}}^4 (\dots (B_{p'_2}^d(\Upsilon_2)) \dots) \right|, \end{aligned}$$

and inequalities (37) and (38). In the last line, we merely took the intersection of the events,  $E_{(\ell_4, \dots, \ell_d)}^c$ , over all  $(\ell_4, \dots, \ell_d)$ , which are independent. When combined, inequalities (34) - (36),

(39), and Lemma 3.1 imply that for  $d \geq 4$ , the probability of  $\mathbf{v}_1$  and  $\mathbf{v}_2$  not being in the same component is at most  $O(n^{-(d+1)})$ . This completes the proof of Lemma 3.4.  $\square$

### 3.4 The Size of the Giant Component

To complete the proof of Theorem 1.2, we need to demonstrate that the proportion of remaining vertices included in the giant component approaches  $(1 - q) > 0$  in probability. To this end we will prove Lemma 3.6, but first we will define  $q$ .

Recall from earlier that  $q_i = \mathbb{P}(\mathbf{Z}_t = \mathbf{0} \text{ for some } t \mid \mathbf{Z}_0 = \mathbf{e}_i)$  is the extinction probability for a multitype branching process in which for any  $k = 1, \dots, d$ ,  $(Z_1^j \mid \mathbf{Z}_0 = \mathbf{e}_k) \sim \text{Poisson}(\lambda a_j)$  if  $j \neq k$  and  $(Z_1^k \mid \mathbf{Z}_0 = \mathbf{e}_k) \equiv 0$ , and initially there is one individual of type  $i$ . The initial vertex in the cluster discovery process gives birth to  $\text{Binomial}(a_i n, \frac{\lambda}{n})$  neighbors of type  $i$  for each  $i$ , and henceforth proceeds like the multitype process in which no vertex can give birth to its own type. The limiting branching process is one in which each binomial birth event is replaced with a Poisson birth event with the same mean. If we consider  $(d - 1)$  independent copies of this Poisson multitype branching process with the modified initial step, and we define  $q$  to be the extinction probability of one of these copies, then the collective process will have the same distribution for all time as the multitype branching process that starts with one individual of each of the  $d$  types. This implies that  $\prod_i q_i = q^{(d-1)}$ . From the theory of multitype branching processes [2], the vector  $(q_1, \dots, q_d)$  is the solution to  $\mathbf{f}(\mathbf{x}) = \mathbf{x}$  for  $\mathbf{x} \in (0, 1)^d$ , where

$$f_i(\mathbf{x}) = \exp \left[ -\lambda \sum_{j \neq i} a_j (1 - x_j) \right].$$

Thus, we have implicitly defined  $q < 1$ .

**Lemma 3.6.**

$$\frac{\#\{\mathbf{v} \in V : |C_{\mathbf{v}}| \leq \beta \log n, \xi_{\mathbf{v}} = 1\}}{\lambda (\prod_i a_i) n^{(d-1)}} \longrightarrow q \quad \text{in probability.} \quad (40)$$

Our approach in the proof of Lemma 3.6 will be to show that the probability of a vertex being included in a component of size at most  $\beta \log n$  given that the vertex is occupied approaches  $q$ . Then we will apply a second moment method argument to demonstrate that the actual proportion of occupied sites in small sized components approaches  $q$ . This will require showing that the events  $\{|C_{\mathbf{v}}| \leq \beta \log n\}$  are asymptotically uncorrelated.

*Proof of Lemma 3.6.* First we will show that  $\mathbb{P}(|C_{\mathbf{v}}| \leq \beta \log n \mid \xi_{\mathbf{v}} = 1) \rightarrow q$ . Let  $\mathbf{Z}_t$  be the multitype branching process with

$$(Z_{t+1}^j \mid \mathbf{Z}_t = \mathbf{e}_k) \begin{cases} \sim \text{Poisson}(\lambda a_j) & j \neq k \\ \equiv 0 & j = k \end{cases}$$

for  $k = 1, \dots, d$  and  $t \geq 1$ , but with  $Z_1^j \sim \text{Poisson}(\lambda a_j)$  for all  $j = 1, \dots, d$ . Then let  $T := 1 + \sum_{t \geq 1} \|\mathbf{Z}_t\|_1$  be the total size of this branching process, and let  $\tau := \inf\{t \mid \|\mathbf{S}_t\|_1 = 0\}$  be the total size of the upper bounding binomial branching process (recall that  $\mathbf{S}_t$  is the random walk version

of this branching process). Also, let the Poisson branching process be coupled with the binomial branching process so as to minimize their total variation distance,  $d_{TV}(Z_1^j, S_1^j) \leq \frac{a_j \lambda}{n}$  [3]. Thus,

$$\mathbb{P}(T \leq \beta \log n, T \neq \tau, \tau \leq \beta \log n) \leq \frac{a_1 \lambda \beta \log n}{n}, \quad (41)$$

$$\mathbb{P}(T \leq \beta \log n, \tau > \beta \log n) \leq \frac{a_1 \lambda \beta \log n}{n}, \quad (42)$$

since in both events the two branching processes must differ in at least one of the first  $\beta \log n$  birth events, and the probability of any birth event differing in the two branching processes is bounded by the total variation distance of their distributions (here we are referring to a randomly distributed number of children of a single type as a ‘birth event’). From the above discussion, we know that  $\mathbb{P}(T < \infty) = q$ . So we have:

$$\begin{aligned} \mathbb{P}(|C_v| \leq \beta \log n \mid \xi_v = 1) &\geq \mathbb{P}(\tau \leq \beta \log n) \\ &\geq \mathbb{P}(T \leq \beta \log n) - 2 \frac{a_1 \lambda \beta \log n}{n} \\ &\rightarrow q, \end{aligned}$$

which implies that  $\liminf \mathbb{P}(|C_v| \leq \beta \log n \mid \xi_v = 1) \geq q$ . For the upper bound, we want to know that up to time  $s = \beta \log n$ , the processes  $\mathcal{A}_t$  and  $\mathcal{S}_t$  are identical with high probability. Let  $T_{\mathcal{A}} := \inf\{t \mid \mathcal{A}_t = \mathbf{0}\}$ . Recall that the number of extra births in  $\mathcal{S}_t$  that are lost in  $\mathcal{A}_t$  up to time  $s \wedge T_{\mathcal{A}}$  can be stochastically bounded above by  $Y^{(s)} \sim \text{Binomial}(d\beta K(\log n)^2, \frac{\lambda}{n})$ , as per (21). Since, for  $n > 2\lambda d\beta K(\log n)^2$ ,

$$\begin{aligned} \mathbb{P}(Y^{(s)} \geq 1) &= \sum_{k=1}^{d\beta K(\log n)^2} \binom{d\beta K(\log n)^2}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{d\beta K(\log n)^2 - k} \\ &\leq \sum_{k=1}^{d\beta K(\log n)^2} \left(\frac{\lambda d\beta K(\log n)^2}{n}\right)^k \\ &\leq \frac{2\lambda d\beta K(\log n)^2}{n}, \end{aligned} \quad (43)$$

the probability that the two processes will differ by time  $s$  is small. Now for the upper bound:

$$\begin{aligned} \mathbb{P}(|C_v| \leq \beta \log n \mid \xi_v = 1) &= \mathbb{P}(\mathcal{A}_s = \mathbf{0} \mid \xi_v = 1) \\ &\leq \mathbb{P}(\mathcal{A}_s = \mathbf{0}, \mathcal{A}_t = \mathcal{S}_t \forall t \leq s \mid \xi_v = 1) \\ &\quad + \mathbb{P}(\mathcal{A}_t \neq \mathcal{S}_t \text{ for some } t \leq s \mid \xi_v = 1) \\ &\leq \mathbb{P}(\mathcal{S}_s = \mathbf{0}, \mathcal{A}_t = \mathcal{S}_t \forall t \leq s \mid \xi_v = 1) + \mathbb{P}(Y^{(s)} \geq 1) \\ &\leq \mathbb{P}(\tau \leq \beta \log n) + \frac{2\lambda d\beta K(\log n)^2}{n} \\ &\leq \mathbb{P}(T \leq \beta \log n) + 2 \frac{a_1 \lambda \beta \log n}{n} + \frac{2\lambda d\beta K(\log n)^2}{n} \\ &\rightarrow q. \end{aligned}$$

The fourth line above uses the estimate in (43) and that  $\mathcal{S}_s$  is independent of the vertex  $v$ , and the fifth line uses the total variation distance bounds that we found earlier for the two branching

processes. Thus we have shown that  $\limsup \mathbb{P}(|C_v| \leq \beta \log n \mid \xi_v = 1) = q$ , and so

$$\lim_{n \rightarrow \infty} \mathbb{P}(|C_v| \leq \beta \log n \mid \xi_v = 1) = q. \quad (44)$$

Now in order to complete the proof, we will apply the second moment method to the following random variables:

$$H_v = \begin{cases} 1 & \text{if } |C_v| \leq \beta \log n, \xi_v = 1 \\ 0 & \text{else.} \end{cases}$$

Note that

$$\begin{aligned} \sum_{v \in V} H_v &= \#\{v \in V : |C_v| \leq \beta \log n, \xi_v = 1\} \\ \sum_{v \in V} \mathbb{E} H_v &= \lambda \left( \prod_i a_i \right) n^{(d-1)} \mathbb{P}(|C_v| \leq \beta \log n \mid \xi_v = 1), \end{aligned}$$

so we wish to prove that

$$\lim_{n \rightarrow \infty} \frac{\sum_{v \in V} H_v}{\sum_{v \in V} \mathbb{E} H_v} = 1$$

in probability. To apply the second moment method we need to bound the variance of the sum of the  $H_v$ 's, and to do this we need an upper bound on the covariance,  $\mathbb{E} H_v H_w - (\mathbb{E} H_v)^2$ . We begin by observing that

$$\mathbb{E} H_v H_w = p^2 \mathbb{P}(|C_v| \leq \beta \log n, |C_w| \leq \beta \log n \mid \xi_v = 1, \xi_w = 1). \quad (45)$$

If  $w \in \mathcal{N}(v)$  then

$$\begin{aligned} \mathbb{P}(|C_v| \leq \beta \log n, |C_w| \leq \beta \log n \mid \xi_v = 1, \xi_w = 1) &= \mathbb{P}(|C_v| \leq \beta \log n \mid \xi_v = 1, \xi_w = 1) \\ &\leq \mathbb{P}(|C_v| \leq \beta \log n \mid \xi_v = 1). \end{aligned}$$

Let us assume now that  $w \notin \mathcal{N}(v)$ , and we will denote by  $J_v \subset V$  such that  $v \in J_v$  an arbitrary set of vertices containing the vertex  $v$ . Then

$$\mathbb{P}(|C_v| \leq \beta \log n, |C_w| \leq \beta \log n \mid \xi_v = 1, \xi_w = 1)$$

$$= \sum_{|J_v| \leq \beta \log n} \sum_{|J_w| \leq \beta \log n} \mathbb{P}(C_v = J_v, C_w = J_w \mid \xi_v = 1, \xi_w = 1)$$

$$= \sum_{|J_v| \leq \beta \log n} \sum_{\substack{|J_w| \leq \beta \log n \\ J_w \cap \mathcal{N}(J_v) = \emptyset}} \mathbb{P}(C_v = J_v, C_w = J_w \mid \xi_v = 1, \xi_w = 1) \quad (46)$$

$$+ \sum_{\substack{|J_v| \leq \beta \log n \\ J_v \ni w}} \mathbb{P}(C_v = J_v = C_w \mid \xi_v = 1, \xi_w = 1), \quad (47)$$

since if  $J_w \cap \mathcal{N}(J_v) \neq \emptyset$  then the event that  $C_w = J_w$  has zero probability unless  $J_w = J_v$ . We will bound lines (46) and (47) separately. Beginning with line (46), we condition on the event that

$C_v = J_v$  to obtain:

$$\begin{aligned} & \sum_{|J_v| \leq \beta \log n} \sum_{\substack{|J_w| \leq \beta \log n \\ J_w \cap \mathcal{N}(J_v) = \emptyset}} \mathbb{P}(C_v = J_v \mid \xi_v = 1, \xi_w = 1) \mathbb{P}(C_w = J_w \mid \xi_v = 1, \xi_w = 1, C_v = J_v) \\ &= \sum_{|J_v| \leq \beta \log n} \mathbb{P}(C_v = J_v \mid \xi_v = 1) \sum_{\substack{|J_w| \leq \beta \log n \\ J_w \cap \mathcal{N}(J_v) = \emptyset}} \mathbb{P}(C_w = J_w \mid \xi_w = 1, C_v = J_v). \end{aligned}$$

The second line is because when  $J_w \cap \mathcal{N}(J_v) = \emptyset$ , the event that  $C_v = J_v$  is independent of  $\xi_w$ , and the event that  $\xi_v = 1$  is contained in the event that  $C_v = J_v$ . Now, if  $C_v = J_v$  and  $J_w \cap \mathcal{N}(J_v) = \emptyset$  then we know that the vertices in  $\mathcal{N}(J_w) \cap \mathcal{N}(J_v)$  must not be occupied, otherwise  $C_v$  would have to include  $J_w$ . Additionally, these are the only vertices that both of the events  $C_v = J_v$  and  $C_w = J_w$  depend upon, otherwise these two events use independent vertices. Noting that  $|\mathcal{N}(J_w) \cap \mathcal{N}(J_v)| \leq 2|J_v||J_w|$ , since each pair of non-neighboring vertices has at most two common neighbors, we have that (46) is bounded by:

$$\begin{aligned} & \leq \sum_{|J_v| \leq \beta \log n} \mathbb{P}(C_v = J_v \mid \xi_v = 1) \sum_{\substack{|J_w| \leq \beta \log n \\ J_w \cap \mathcal{N}(J_v) = \emptyset}} \frac{\mathbb{P}(C_w = J_w \mid \xi_w = 1)}{\left(1 - \frac{\lambda}{n}\right)^{2\beta^2(\log n)^2}} \\ & \leq \left(1 + \frac{5\lambda\beta^2(\log n)^2}{n}\right) \sum_{|J_v| \leq \beta \log n} \mathbb{P}(C_v = J_v \mid \xi_v = 1) \sum_{|J_w| \leq \beta \log n} \mathbb{P}(C_w = J_w \mid \xi_w = 1) \\ & = \left(1 + \frac{5\lambda\beta^2(\log n)^2}{n}\right) \frac{(\mathbb{E}H_v)^2}{p^2}. \end{aligned} \tag{48}$$

In the first line above we used that the event that none of the vertices in  $\mathcal{N}(J_w) \cap \mathcal{N}(J_v)$  are occupied is contained in the event that  $C_w = J_w$ . In the second line we applied the following computation, in which the inequalities hold provided  $n \geq 2\lambda$  and  $n \geq 32\lambda\beta^2(\log n)^2$ :

$$\begin{aligned} \left(\frac{1}{1 - \frac{\lambda}{n}}\right)^{2\beta^2(\log n)^2} & \leq \left(1 + \frac{2\lambda}{n}\right)^{2\beta^2(\log n)^2} \\ & \leq 1 + \frac{4\lambda\beta^2(\log n)^2}{n} + \frac{32\lambda^2\beta^4(\log n)^4}{n^2} \\ & \leq 1 + \frac{5\lambda\beta^2(\log n)^2}{n}. \end{aligned}$$

To bound line (47), we observe that

$$\begin{aligned} & \sum_{\substack{|J_v| \leq \beta \log n \\ J_v \ni w}} \mathbb{P}(C_v = J_v = C_w \mid \xi_v = 1, \xi_w = 1) = \mathbb{P}(|C_v| \leq \beta \log n, C_v = C_w \mid \xi_v = 1, \xi_w = 1) \\ & \leq \mathbb{P}(A_t \cap \mathcal{N}(w) \neq \emptyset \text{ for some } t \leq \beta \log n \mid A_0 = \{v\}) \\ & \leq 1 - \left(1 - \frac{\lambda}{n}\right)^{2\beta \log n} \\ & \leq \frac{2\lambda\beta \log n}{n} + \frac{8\lambda^2\beta^2(\log n)^2}{n^2} \\ & \leq \frac{3\lambda\beta \log n}{n} \end{aligned} \tag{49}$$

In words, the second line above says that the probability that  $C_v$  has fewer than  $\beta \log n$  vertices and contains the vertex  $w$  (conditional on  $v$  and  $w$  begin occupied) is at most the probability that we find an occupied neighbor of  $w$  by time  $\beta \log n$  during the process of discovering the cluster  $C_v$ . In the third line, we used the fact that at each time,  $t \leq \beta \log n$ , until  $A_t \cap \mathcal{N}(w) \neq \emptyset$  we have that  $|\mathcal{N}(v_t) \cap \mathcal{N}(w)| \leq 2$ , and at least one such vertex must be occupied in order for  $w$  to be discovered. The last two lines require that  $n \geq 8\lambda\beta \log n$ .

Combining (45) – (49) shows that the covariance between  $H_v$  and  $H_w$  when  $w \notin \mathcal{N}(v)$  can be bounded as:

$$\begin{aligned} \mathbb{E}H_v H_w - (\mathbb{E}H_v)^2 &\leq \frac{5\lambda\beta^2(\log n)^2}{n}(\mathbb{E}H_v)^2 + \frac{3\lambda^3\beta \log n}{n^3} \\ &\leq \frac{6\lambda^3\beta^2(\log n)^2}{n^3} \end{aligned} \quad (50)$$

for sufficiently large  $n$ , since  $\mathbb{E}H_v \leq p$ . Now we can use this to bound the variance:

$$\begin{aligned} \text{Var} \left( \sum_{v \in V} H_v \right) &= \sum_{\substack{v, w \in V \\ w \in \mathcal{N}(v)}} \mathbb{E}H_v H_w + \sum_{\substack{v, w \in V \\ w \notin \mathcal{N}(v)}} \mathbb{E}H_v H_w - \sum_{v, w \in V} \mathbb{E}H_v \mathbb{E}H_w \\ &\leq \sum_{\substack{v, w \in V \\ w \in \mathcal{N}(v)}} \mathbb{E}H_v H_w + \sum_{\substack{v, w \in V \\ w \notin \mathcal{N}(v)}} [\mathbb{E}H_v H_w - \mathbb{E}H_v \mathbb{E}H_w] \\ &\leq \lambda^2 \left( \prod_i a_i \right) a_1 d n^{(d-1)} + 6\lambda^3\beta^2 \left( \prod_i a_i \right)^2 n^{(2d-3)} (\log n)^2 \\ &= O \left( n^{(2d-3)} (\log n)^2 \right). \end{aligned} \quad (51)$$

Using this bound on the variance and Chebyshev's inequality yields:

$$\begin{aligned} \mathbb{P} \left( \left| \frac{\sum_{v \in V} H_v}{\sum_{v \in V} \mathbb{E}H_v} - 1 \right| > n^{-1/3} \right) \\ &= \mathbb{P} \left( \left| \sum_{v \in V} H_v - \sum_{v \in V} \mathbb{E}H_v \right| > \lambda \left( \prod_i a_i \right) n^{(d-1)} \mathbb{P}(|C_v| \leq \beta \log n \mid \xi_v = 1) n^{-1/3} \right) \\ &\leq \frac{\text{Var}(\sum_{v \in V} H_v)}{\lambda^2 (\prod_i a_i)^2 (q + o(1))^2 n^{(2d-2-2/3)}} \\ &= O \left( n^{-1/3} (\log n)^2 \right). \end{aligned} \quad (52)$$

Combining (44) and (52) completes the proof of Lemma 3.6.  $\square$

Thus the proof of Theorem 1.2 is complete.  $\square$

## 4 Connectivity Threshold

We have demonstrated that when  $p = \lambda/n$ , a giant component emerges at  $\lambda_c$ . The purpose of this section is to determine when the random site subgraph of the Hamming torus is connected. We prove the following theorems.

**Theorem (1.4 restated).** Let  $c < \frac{d-1}{\sum a_i}$ . If  $p = p(n) \leq \frac{c \log n}{n}$  and  $p = \omega(n^{-d})$  then the random site subgraph of the Hamming torus contains isolated vertices, and is thus not connected (a.a.s.).

**Theorem (1.5 restated).** Let  $c > \frac{d-1}{\sum a_i}$ . If  $p = p(n) \geq \frac{c \log n}{n}$  then the random site subgraph of the Hamming torus is connected (a.a.s.).

**Theorem (1.6 restated).** Fix  $a_1 \geq a_2 \geq \dots \geq a_d$ , and let  $c > \frac{d-1}{2 \sum_{i=2}^d a_i + a_1}$ . If  $p = \frac{c \log n}{n}$  then every vertex in the random site subgraph of the Hamming torus is either isolated or belongs to the giant component (a.a.s.).

These three theorems together imply that, with probability approaching one, the random site subgraph of the Hamming torus is connected if and only if it contains no isolated vertices (except in the trivial case  $p \asymp n^{-d}$ , where the subgraph may consist of just a single vertex with positive probability).

*Proof of Theorem 1.4.* We denote by  $d_v$  the degree of the vertex  $v$  in the random site subgraph of the Hamming torus, with the convention that  $d_v = -1$  if  $\xi_v = 0$ . Let  $I_n$  be the number of isolated vertices in the random site subgraph. Then

$$\begin{aligned} \mathbb{E}I_n &= p \sum_{v \in V} \mathbb{P}(d_v = 0 \mid \xi_v = 1) \\ &= p \left( \prod a_i \right) n^d (1-p)^{\sum a_i n - d} \\ &\asymp p n^d \exp \left[ -p \sum a_i n \right], \end{aligned} \tag{53}$$

where “ $a_n \asymp b_n$ ” means that  $0 < \lim \frac{a_n}{b_n} < \infty$ . The last line above follows from the limit

$$\lim_{n \rightarrow \infty} (1-p)^n e^{pn} = 1$$

provided  $p = o(n^{-1/2})$ . If  $n^{-1} \leq p \leq \frac{c \log n}{n}$ , then equation (53) is bounded by

$$\mathbb{E}I_n \gtrsim n^{d-1-c \sum a_i} \rightarrow \infty.$$

Otherwise, if  $n^{-d+1} \ll p \leq n^{-1}$ , then equation (53) yields

$$\mathbb{E}I_n \gtrsim p n^{d-1} e^{-\sum a_i} \rightarrow \infty.$$

We will now use the second moment method to show that  $I_n$  stays close to its mean.

$$\begin{aligned}
\mathbb{E}I_n^2 &= \sum_{\mathbf{v} \in V} \sum_{\mathbf{w} \in V} \mathbb{P}(d_{\mathbf{v}} = 0, d_{\mathbf{w}} = 0) \\
&= \sum_{\mathbf{v} \in V} \mathbb{P}(d_{\mathbf{v}} = 0) + \sum_{\substack{\mathbf{v}, \mathbf{w} \in V \\ d(\mathbf{v}, \mathbf{w})=1}} \mathbb{P}(d_{\mathbf{v}} = 0, d_{\mathbf{w}} = 0) \\
&\quad + \sum_{\substack{\mathbf{v}, \mathbf{w} \in V \\ d(\mathbf{v}, \mathbf{w})=2}} \mathbb{P}(d_{\mathbf{v}} = 0, d_{\mathbf{w}} = 0) + \sum_{\substack{\mathbf{v}, \mathbf{w} \in V \\ d(\mathbf{v}, \mathbf{w}) \geq 3}} \mathbb{P}(d_{\mathbf{v}} = 0, d_{\mathbf{w}} = 0) \\
&= \mathbb{E}I_n + 0 + \sum_{\substack{\mathbf{v}, \mathbf{w} \in V \\ d(\mathbf{v}, \mathbf{w})=2}} \mathbb{P}(d_{\mathbf{v}} = 0 \mid d_{\mathbf{w}} = 0) \mathbb{P}(d_{\mathbf{w}} = 0) + \sum_{\substack{\mathbf{v}, \mathbf{w} \in V \\ d(\mathbf{v}, \mathbf{w}) \geq 3}} \mathbb{P}(d_{\mathbf{v}} = 0) \mathbb{P}(d_{\mathbf{w}} = 0) \\
&= \mathbb{E}I_n + \sum_{\substack{\mathbf{v}, \mathbf{w} \in V \\ d(\mathbf{v}, \mathbf{w})=2}} \frac{\mathbb{P}(d_{\mathbf{v}} = 0) \mathbb{P}(d_{\mathbf{w}} = 0)}{(1-p)^2} + \sum_{\substack{\mathbf{v}, \mathbf{w} \in V \\ d(\mathbf{v}, \mathbf{w}) \geq 3}} \mathbb{P}(d_{\mathbf{v}} = 0) \mathbb{P}(d_{\mathbf{w}} = 0)
\end{aligned}$$

In the second line above, the second sum is equal to 0 because if  $\mathbf{v}$  and  $\mathbf{w}$  are neighbors, then they cannot both be isolated. The last sum in the second line is equal to the last sum in the third line because if  $d(\mathbf{v}, \mathbf{w}) \geq 3$ , then the events  $\{d_{\mathbf{v}} = 0\}$  and  $\{d_{\mathbf{w}} = 0\}$  rely on disjoint sets of vertices and are thus independent. In the case where  $d(\mathbf{v}, \mathbf{w}) = 2$ , these events share exactly two vertices in common, thus  $\mathbb{P}(d_{\mathbf{v}} = 0) = \mathbb{P}(d_{\mathbf{v}} = 0 \mid d_{\mathbf{w}} = 0) (1-p)^2$ . Let  $N_k = \#\{\mathbf{v} \in V \mid d((1, 1, \dots, 1), \mathbf{v}) = k\}$  be the number of neighbors at exactly Hamming distance  $k$  that a vertex in the Hamming torus has. Then

$$\begin{aligned}
\mathbb{E}I_n^2 &= \mathbb{E}I_n + |V| N_2 \frac{\mathbb{P}(d_{\mathbf{v}} = 0)^2}{(1-p)^2} + |V| (|V| - N_2 - N_1 - N_0) \mathbb{P}(d_{\mathbf{v}} = 0)^2 \\
&\lesssim \mathbb{E}I_n + |V| N_2 \mathbb{P}(d_{\mathbf{v}} = 0)^2 (1+5p) + |V| (|V| - N_2) \mathbb{P}(d_{\mathbf{v}} = 0)^2 \\
&= \mathbb{E}I_n + 5p |V| N_2 \mathbb{P}(d_{\mathbf{v}} = 0)^2 + |V|^2 \mathbb{P}(d_{\mathbf{v}} = 0)^2 \\
&\leq (1 + O(\log^2 n)) \mathbb{E}I_n + (\mathbb{E}I_n)^2,
\end{aligned}$$

$$\text{Var}(I_n) = O(\log^2 n) \mathbb{E}I_n. \quad (54)$$

In the second line above we applied the estimate  $(1-p)^{-2} \leq (1+2p)^2 \leq (1+5p)$  whenever  $p < 1/4$ . In the fourth line we used the following facts:  $\mathbb{E}I_n = |V| \mathbb{P}(d_{\mathbf{v}} = 0)$ ,  $\mathbb{P}(d_{\mathbf{v}} = 0) \leq p \leq \frac{c \log n}{n}$ , and  $N_2 = O(n^2)$ . We can now bound the fluctuations in  $I_n$  by

$$\mathbb{P}\left(|I_n - \mathbb{E}I_n| > (\mathbb{E}I_n)^{1/2} \log^2 n\right) \leq \frac{\text{Var}(I_n)}{\mathbb{E}I_n \log^4 n} = O((\log n)^{-2}). \quad (55)$$

Since  $\mathbb{E}I_n \gg n^\epsilon \gg \log^4 n$  for some  $\epsilon > 0$ , this completes the proof of Theorem 1.4.  $\square$

The proofs for Theorems 1.5 and 1.6 will be similar in that they appeal to the machinery that we have already developed in the proof of Theorem 1.2. In particular, we will merely replace Lemma 3.3 with a more assertive statement, then apply Lemmas 3.1, 3.4 and the arguments of Section 3.1 with little modification.

*Proof of Theorem 1.5.* Let  $p = \frac{c \log n}{n}$ , and observe that monotonicity implies the result for larger  $p$ . Recall how we coupled a lower bounding random walk,  $\mathbf{W}_t$ , with  $A'_t$  in Section 3.1. We wish to construct the same lower bounding walk here, and we will even use the same parameter  $p_1 = \frac{\lambda_1}{n}$  where  $\lambda_1$  is a constant such that  $\lambda_c < \lambda_1 < \lambda_c^{(d-1)}$ . The only difference here is that now  $p = p_1 + p_2 - p_1 p_2$  where  $p_2 \geq \frac{c \log n}{n}$ , which is okay because all we need to apply Lemma 3.4 is that  $p_2 \geq \frac{c}{n}$ . We then can apply Lemma 3.1 and repeat the arguments at the end of Section 3.1 to again obtain equation (13), which says that, provided  $|A'_0| \geq m = \frac{d+1}{\gamma_2} \log n$ ,  $|A'_0| = o(n^{2/3})$  and  $|R'_0| = o(n^{2/3})$ , the process of discovering occupied vertices with parameter  $p_1$  will survive until time  $r = n^{d-4/3}$ . At this point, we can apply Lemma 3.4 to say that with high probability every cluster reaching size  $r$  will connect into a single giant component. Thus, to show that the entire subgraph is connected, we need to show that either  $|C_v| \geq m$  or  $|C_v| = 0$  for every  $v \in V$  with probability approaching 1.

Consider a fixed vertex,  $v$ . Disregarding prior notation for the moment, let  $A_k$  be the set of vertices in the random site subgraph of the Hamming torus to which there exists a unique shortest path of length  $k$  from  $v$ . Then for any constant  $\ell$

$$\begin{aligned}
\mathbb{P}_p(|A_2| < m \mid |A_1| = \ell) &= \mathbb{P}_p\left(e^{-|A_2|} > e^{-m} \mid |A_1| = \ell\right) \\
&\leq e^m \mathbb{E}_p\left[e^{-|A_2|} \mid |A_1| = \ell\right] \\
&= e^m \left[1 - p(1 - e^{-1})\right]^{|\mathcal{N}(A_1) \setminus \mathcal{N}(v)|} \\
&\leq \exp\left[m - p(1 - e^{-1})|\mathcal{N}(A_1) \setminus \mathcal{N}(v)|\right] \\
&\leq \exp\left[m - \frac{c \log n}{n}(1 - e^{-1})\left(\sum_{i=2}^d a_i n - \ell - d + 1\right)\ell\right] \\
&= \exp\left[\left(\frac{d+1}{\gamma_2} - c(1 - e^{-1})\sum_{i=2}^d a_i\right)\log n + c(1 - e^{-1})\ell(\ell + d - 1)\frac{\log n}{n}\right].
\end{aligned} \tag{56}$$

In the fourth line above we used the bound  $(1 - x) \leq e^{-x}$ . In the fifth line above we assumed WLOG that  $a_1 \geq a_2 \geq \dots \geq a_d$ , and used the fact that each of the  $\ell$  vertices in  $A_1$  has at least  $(a_2 + \dots + a_d)n - \ell - d + 1$  neighbors that are not also neighbors of  $v$  or any other vertex in  $A_1$ . Now, if we choose  $\ell$  to be any constant so that

$$\begin{aligned}
\frac{d+1}{\gamma_2} - \ell c(1 - e^{-1})\sum_{i=2}^d a_i &\leq -(d+1) \\
\ell &\geq \frac{(d+1)(1 + 1/\gamma_2)}{c(1 - e^{-1})\sum_{i=2}^d a_i},
\end{aligned} \tag{57}$$

then (56) implies that

$$\mathbb{P}_p(|A_2| < m \mid |A_1| = \ell) = O\left(n^{-(d+1)}\right). \tag{58}$$

Let  $\mathbf{Z}_t$  be the multitype branching process with  $Z_1^i \sim \text{Binomial}(a_i n, p)$  for all  $i = 1, \dots, d$ , and for all  $t \geq 1$ ,  $(Z_{t+1}^j \mid \mathbf{Z}_t = \mathbf{e}_i) \sim \text{Binomial}(a_j n, p)$  for all  $i$  and  $j \neq i$ , and  $(Z_{t+1}^i \mid \mathbf{Z}_t = \mathbf{e}_i) \equiv 0$ . Observe

that  $|A_k| \leq \|\mathbf{Z}_k\|_1$  by a now familiar coupling. For  $\theta = \log 5 \approx 1.6$

$$\begin{aligned}
\mathbb{P}\left(\|\mathbf{Z}_1\|_1 \geq 5 \sum a_i c \log n\right) &\leq e^{-5\theta \sum a_i c \log n} \mathbb{E} e^{\theta \|\mathbf{Z}_1\|_1} \\
&\leq e^{-5\theta \sum a_i c \log n} [1 + p(e^\theta - 1)]^{\sum a_i n} \\
&\leq \exp\left[-5\theta \sum a_i c \log n + (e^\theta - 1) \sum a_i c \log n\right] \\
&\leq \exp\left[(e^\theta - 1 - 5\theta) \sum a_i c \log n\right] \\
&\leq \exp[-4(d-1) \log n] \\
&= O(n^{-(d+2)}).
\end{aligned} \tag{59}$$

Now, applying a union bound,

$$\begin{aligned}
\mathbb{P}\left(\|\mathbf{Z}_2\|_1 \geq \left(5 \sum a_i c \log n\right)^2 \mid \|\mathbf{Z}_1\|_1 \leq 5 \sum a_i c \log n\right) \\
\leq \left(5 \sum a_i c \log n\right) \mathbb{P}\left(\|\mathbf{Z}_1\|_1 \geq 5 \sum a_i c \log n\right) \\
= O(n^{-(d+1)}).
\end{aligned} \tag{60}$$

Combining the last two inequalities shows that  $|A_1|$  and  $|A_2|$  are both  $o(n^{2/3})$  with high probability. So we can apply Lemma 3.1 with  $A'_0 = A_2$ ,  $R'_0 = A_1 \cup \{\mathbf{v}\}$  and  $U'_0 = V \setminus \mathcal{N}(R'_0)$ , and continue the arguments discussed above to show that all vertices for which  $d_{\mathbf{v}} \geq \ell$  will be in the same large component with probability exceeding  $1 - O(n^{-1})$  (by a union bound over all vertices in  $V$ ). Thus, we need to show that all vertices either have  $d_{\mathbf{v}} = -1$  or  $d_{\mathbf{v}} \geq \ell$ .

$$\begin{aligned}
\mathbb{P}(d_{\mathbf{v}} < \ell \mid d_{\mathbf{v}} \neq -1) &= \sum_{j=0}^{\ell-1} \binom{\sum a_i n - d}{j} \left(\frac{c \log n}{n}\right)^j \left(1 - \frac{c \log n}{n}\right)^{\sum a_i n - d - j} \\
&\leq \sum_{j=0}^{\ell-1} \frac{(c \log n)^j}{j!} \left(\sum a_i\right)^j e^{-\sum a_i c \log n + (d+j)c \log n/n} \\
&\leq \ell (c \log n)^{\ell-1} n^{-\sum a_i c} e^{(d+\ell-1)c \log n/n} \max\left\{1, \left(\sum a_i\right)^{\ell-1}\right\} \\
&= O(n^{-\sum a_i c} (\log n)^{\ell-1}).
\end{aligned}$$

So, since  $\sum a_i c > d - 1$ , we have that

$$\mathbb{P}(0 \leq d_{\mathbf{v}} < \ell) = O(n^{-\sum a_i c - 1} (\log n)^\ell) = o(n^{-d}),$$

and by the union bound

$$\mathbb{P}(0 \leq d_{\mathbf{v}} < \ell \text{ for some } \mathbf{v} \in V) = o(1).$$

This completes the proof of Theorem 1.5.  $\square$

*Proof of Theorem 1.6.* The proof here will be very similar to that of Theorem 1.5, except that here we will show that with probability approaching 1 one of three things can occur for each  $\mathbf{v} \in V$ :  $|C_{\mathbf{v}}| = 0$ ,  $|C_{\mathbf{v}}| = 1$  or  $|C_{\mathbf{v}}| \geq m$ . The last case implies, by the discussion at the start of the last proof, that  $\mathbf{v}$  is in the giant component. We will also rely on the same notation from the last proof.

If  $d_v \leq 0$  then  $v$  is either isolated or not in the subgraph, so we consider the cases where  $d_v \geq 1$ . If  $d_v \geq \ell$ , where  $\ell$  is defined according to (57), then equations (56) and (58) still hold, as well as the arguments that follow, so  $v$  will be in the giant component with probability approaching one. Now we must consider the case where  $1 \leq d_v < \ell$ . Suppose  $d_v = l$  where  $1 \leq l < \ell$ , then

$$\begin{aligned}
& \mathbb{P}(|A_2| < \ell, d_v = l) \\
&= p \binom{\sum a_i n}{l} p^l (1-p)^{\sum a_i n - d - l} \sum_{j=0}^{\ell-1} \binom{|\mathcal{N}(A_1) \setminus \mathcal{N}(v)|}{j} p^j (1-p)^{|\mathcal{N}(A_1) \setminus \mathcal{N}(v)| - j} \\
&\leq \sum_{j=0}^{\ell-1} \left( l \sum a_i \right)^{j+l} \frac{(c \log n)^{j+l+1}}{j!} n^{-c((l+1) \sum_{i=2}^d a_i + a_1) - 1} e^{(j+l(l+d-1))c \log n/n} \\
&\leq \ell \max \left\{ \left( \sum a_i \right)^l, \left( \sum a_i \right)^{\ell+l} \right\} (c \log n)^{\ell+1} n^{-c((l+1) \sum_{i=2}^d a_i + a_1) - 1} e^{(\ell+l(l+d-1))c \log n/n} \\
&= o(n^{-d}). \tag{61}
\end{aligned}$$

The second line above is because each of the  $l$  vertices in  $A_1$  have at least  $(a_2 + \dots + a_d)n - l - d + 1$  neighbors that are not also neighbors of  $v$  or any other vertex in  $A_1$ . The last line above is because  $c > \frac{d-1}{2 \sum_{i=2}^d a_i + a_1}$ . Summing the above probabilities implies that  $\mathbb{P}(|A_2| < \ell, 1 \leq d_v < \ell) = o(n^{-d})$ . So for every  $v \in V$ , if  $d_v \leq 0$ , then  $v$  is either isolated or not in the random subgraph. If  $d_v \geq \ell$ , then  $v$  is in the giant component with probability approaching one. The probability of none of the last three cases occurring and  $v$  having fewer than  $\ell$  occupied vertices at distance 2 approaches 0. To complete the proof, we merely need to show that if  $v$  has  $\ell$  occupied vertices at distance 2 then  $v$  will have at least  $m$  vertices at distance 3. This is easy because the arguments are identical to those of inequalities (56), (58)-(60), but with  $|A_2|$  and  $|A_1|$  replaced by  $|A_3|$  and  $|A_2|$ , and  $\ell$  replaced by  $4\ell$  in inequality (56). This last substitution is because if  $|A_1| < \ell$  and  $|A_2| = \ell$  then each vertex in  $A_2$  has at least  $(a_2 + \dots + a_d)n - 4\ell - d + 1$  neighbors that are not also neighbors of  $v$  or any other vertex in  $A_1$  or  $A_2$ . This completes the proof of Theorem 1.6.  $\square$

## References

- [1] M. Ajtai, J. Komlós, and E. Szemerédi. Largest random component of a k-cube. *Combinatorica*, 2(1):1–7, 1981.
- [2] K. B. Athreya and P. E. Ney. *Branching Processes*. Dover Publications, Inc., 1972.
- [3] A. D. Barbour, L. Holst, and S. Janson. *Poisson Approximation*. Oxford University Press, 1992.
- [4] B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms*, 31(1):3–122, 2007.
- [5] B. Bollobás, Y. Kohayakawa, and T. Łuczak. On the evolution of random boolean functions. *Extremal Problems for Finite Sets, Visegrád (Hungary)*, pages 137–156, 1991.
- [6] C. Borgs, J. T. Chayes, R. van der Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs: I. The scaling window under the triangle condition. *Random Structures and Algorithms*, 27:137–184, 2005.

- [7] C. Borgs, J. T. Chayes, R. van der Hofstad, G. Slade, and J. Spencer. Random subgraphs of finite graphs: II. The lace expansion and the triangle condition. *Annals of Probability*, 33:1886–1944, 2005.
- [8] R. Durrett. *Random Graph Dynamics*. Cambridge University Press, 2007.
- [9] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 5:17–61, 1960.
- [10] S. Gavrillets. *Fitness Landscapes and the Origin of Species*. Princeton University Press, 2004.
- [11] A. J. Schwenk and J. I. Munro. How small can the mean shadow of a set be? *The American Mathematical Monthly*, 90(5):325–329, 1983.
- [12] R. van der Hofstad and M. Luczak. Random subgraphs of the 2D Hamming graph: the supercritical phase. *Probability Theory and Related Fields*, 2009. <http://www.springerlink.com/content/f373823235478843/>.